Department of Electrical and Computer Engineering

Statistical and Probabilistic Models for Smart Electricity Distribution Networks

Yingliang Li

This thesis is presented for the Degree of
Doctor of Philosophy

of

Curtin University

July 2013

DECLARATION

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: …………………………………
(Yingliang Li)

Date: ………………………........................

ABSTRACT

This thesis is aimed towards an improved statistical understanding of distribution feeders, an improved probabilistic understanding of their loads and methods for network-wide assessment of Smart Grid technologies.

The scale and complexity of the distribution network makes it practically impossible to fully analyse the impact of a new technology in each and every corner of the network. A reasonable approach is to develop representative test cases, for both networks and consumer load to make the technology assessment process practicable.

A compact but statistically representative test set allows distribution researchers and engineers to quickly evaluate smart grid technologies with confidence that the results will be meaningful in the physical network.

There has been an international recognition of the need for representative test cases. MV test feeders have been provided by organisations, such as the Pacific Northwest National Laboratory (PNNL) of Department of Energy (DOE), the United Kingdom Generic Distribution System (UKGDS) and IEEE.

Australian engineers, software programmers and researchers need feeder models for the same reasons they are needed internationally to test technologies, especially important for MV level technologies, such as switched capacitor, integrated volt/VAR controller (IVVC), larger distributed generation (DG), dSTATCOMS. Overseas feeder MV models, especially IEEE test feeders, have been frequently used in Australia. Some models are reasonable, but others, such as IEEE 13 bus system which used 4.16kV branches and components, are irrelevant to the Australian setting. As for the LV models, there is a critical need. There are no published representative models. If models were developed there are specific geographical peculiarities. The US LV system is completely different from Australia.

This thesis presents a highly efficient multivariable statistical analysis method that relies upon a few key variables which are highly meaningful from an engineering perspective and readily available in most if not all distribution companies. The method

attempts to obtain a stable classification results with minimum error rate. A set of discriminant functions has been extracted as feeder classifiers. The work identified statistically significant feeders with a limited number of engineering variables using two very different data bases – a MV feeder data base of 204 members and an LV data base of 8,858 members. The capability of the general method to form useful cluster classifications in such disparate applications illustrates the universality of the method.

In the LV system, there are limited models for consumer load. Most utilities develop a "deemed profile" for residential customers and this is normally determined by the observation of group demands on distribution transformers or medium voltage feeders with a dominance of residential load. The LV networks are designed to provide adequate voltage regulation at the periods of peak demand and the design is often based on a single agreed demand figure for a residential consumer. These approaches cannot incorporate a more detailed understanding of consumer load across the diurnal cycle and have led to design approaches that contribute to power quality problems such as voltage rise in systems with high penetrations of PV. Therefore, this thesis seeks to develop a much more nuanced understanding of consumer load and to present models that can support the analysis of new technologies.

Hybrid models have been built for high and low demand days modelling of residential consumer load. The model for underlying diurnal variation has been constructed using a truncated Fourier series representation. Cluster and discriminant analysis have been used to extract six representative daily load profiles. After the low frequency variation being established, a high frequency residual remains. This has been modelled using a power spectrum. The load models have been validated using the data from the test days. For each test day load ensembles were constructed that included appropriate numbers of residential loads of each cluster type. The load groups successfully produced aggregated load behaviours that compare well with the physical load data.

The outcomes of the thesis will assist the distribution network operators in their technical decision making and better understanding in the deployment of Smart Grid technologies. The analysis and modeling methods developed are transferrable to other networks and load groups.

To my wife Jing

and

my baby Junchen

ACKNOWLEDGMENT

First and foremost, I want to express my greatest gratitude to my Ph.D. supervisor Professor Peter J. Wolfs. He is such a wonderful advisor and mentor. Under his guidance, I have learned a lot in different aspects of conducting research. I thank him for every piece of his intensive efforts that have been put into this research work. I also appreciate that he continued to supervise me and meet me online every week till submission of this thesis even though he relocated back to Central Queensland University at the end of my Ph.D. He is such a great person, and I am grateful for everything he has done for me.

I am thankful to another supervisor of mine - Dr. Dilan Jayaweera for his supportive guidance and advice. I also want to thank him for his kind support during my thesis writing. I wish to express my gratitude to Professor Syed Islam for giving valuable sources of knowledge and suggestions during my Ph.D..

I'd like to thank Dr. Yee-Hong Leung and Professor Mohammad Masoum for their great efforts in providing such a good environment for research in the department. I also want to express my thanks to Patricia Tang and Margaret Pittuck for their kind assistance. I am grateful to the Ph.D. students in my office who have helped me during my study. Their emotional support and friendship is something I shall never forget.

I would like to express my thanks to the support of Western Power Corporation for providing feeder data, technical reports and photographs. I especially thank Peter Martino, Ishaan Khanna, Sundy Tjhin and Oscar Arteaga for their kind help in identifying, explaining and extracting data sets.

Let me reserve my final appreciation to my family. Without the care and love from my father and mother, I definitely could not have completed my doctoral degree. They have always offered me so much and asked me for so little. Special thanks to my wife, whose sincerest love has given me courage to face difficulties in my way of pursuing my dream. I am very proud of my son, who newly joined the family. He has been given me huge pleasure and kept me so energised throughout the final stage of my Ph.D. study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

c_i	Smaller disjoint sets of feeder data after clustering
df	Degree of freedom
F	F ratio measuring the level of significance of a variable
g	Number of cells with non-singular covariance matrices
m_{opt}	Optimal number of clusters
n	Total sample size
n_i	Number of cases in the i_{th} cell.
P_n	Demand of customer at the time of system maximum demand
S	Pooled covariance matrix
S_i	Cell covariance matrix
SS_{bg}	Cross-products matrices between-group differences
SS_{wg}	Cross-products matrices within-groups differences
SSE	Sum of squares for error
SSE_k	Internal cluster distance for cluster k
T	Total distance for clusters
TSS	Total sum of squares
u_j	Clustering input variables
WSS	Within group sum of squares
Z	Feeder data composed of vectors z_i
ϕ_r	Spectral density
λ_i	Ordered eigenvalues of $SS_{wg}^{-1}SS_{bg}$

Variables for distribution feeder classification

L_{UG}	Underground feeder length
L_{OH}	Overhead feeder length
R	Total load capacity/rating ratio
P_{others}	Load capacity excluding residential customer
N_{others}	Number of customers excluding residential customers,

	which includes number of commercial, industrial and 24/7-hour
P_r	Load capacity of residential customer
N_r	Number of residential customer
P_L	Peak load
N_C	Number of customers
L_{MV}	MV circuit length
L_{LV}	LV circuit length
N_{TX}	Customers per transformer
C_{LV}	Installed capacity per customer

LIST OF PUBLICATIONS

[1] Yingliang Li, Peter J. Wolfs, A taxonomic description for Western Australian distribution MV and LV feeders, IET Generation, Transmission & Distribution. (To be published, accepted on 11th May 2013, GTD-2013-0005, Digital Object Identifier: 10.1049/iet-gtd.2013.0005)

[2] Yingliang Li, Peter J. Wolfs, ‘A hybrid model for residential loads in a distribution system with high PV penetration’, Power Systems, IEEE Transactions on, vol. 28, no. 3, pp. 3372-3379, 2013.

[3] Yingliang Li, Peter Wolfs, Statistical Identification of Prototypical Low Voltage Distribution Feeders in Western Australia, in Power and Energy Society General Meeting, IEEE, 2012.

[4] Yingliang Li, Peter Wolfs, Statistical discriminant analysis of high voltage feeders in Western Australia distribution networks, in Power and Energy Society General Meeting, IEEE, 2011.

[5] Yingliang Li, Peter Wolfs, Preliminary statistical study of low voltage distribution feeders under a representative HV network in Western Australia, in Universities Power Engineering Conference (AUPEC), 21st Australasian, 2011.

[6] Yingliang Li, Peter Wolfs, A statistical study on topological features of high voltage distribution networks in Western Australia, in Universities Power Engineering Conference (AUPEC), 20th Australasian, 2010.

Chapter 1 Introduction

1.1 Research background

This thesis focuses on necessary improvements to the modelling of distribution networks and the modelling of load to support utility engineers as they strive to design and implement the Smart Grid. The Smart Grid is an overarching term that captures the application of modern communications and control technologies to modernise the power grid [1].

The conventional power grid is being challenged in many respects. It has been experiencing difficulties in catching up with new and increasing demands, and is outdated. Smart Grid is an auto-regulating, self-monitoring power grid that accepts electrical generation from any source of fuel (coal, sun, wind) and delivers that energy to satisfy a consumer's end use with minimal human intervention. The intention of Smart Grid is to modernise the grid infrastructure, build in intelligence to power grids, deliver economic, secure, and efficient systems to customer premises. [2-10] A series of new technologies are gradually making this vision into reality.

In the near future, Smart Grid will enable electricity and information flowing in real time, near-zero economic losses from outages and power quality disturbances, a wider array of customised energy choices, suppliers competing in open markets to provide the world's best electric services. Customers could expect to routinely obtain electricity services at improved reliability and quality levels tailored to their needs with greatly reduced environmental impacts. All of the vision will be supported by a new energy infrastructure built on superconductivity, distributed intelligence and resources, clean power, and the hydrogen economy [11].

Internationally the Smart Grid agenda has been driven by a number of national plans. In the USA the "Grid 2030" vision [11], started a chain of reforms that was intensively driven by the Obama Administration partially to spur economic growth after the Global Financial Crisis (GFC). One of the most comprehensive and best accepted models for the Smart Grid, the IEEE conceptual model, was a consequence of this activity [12] as seen in Figure 1.1. This model includes seven domains - bulk generation, transmission, distribution, customers, operations, markets and service providers. Each domain has Smart Grid elements that are connected to each other through two-way communications and energy paths. In Europe, Smart Grid development was guided by the European Platform Strategic Research Agenda [13].

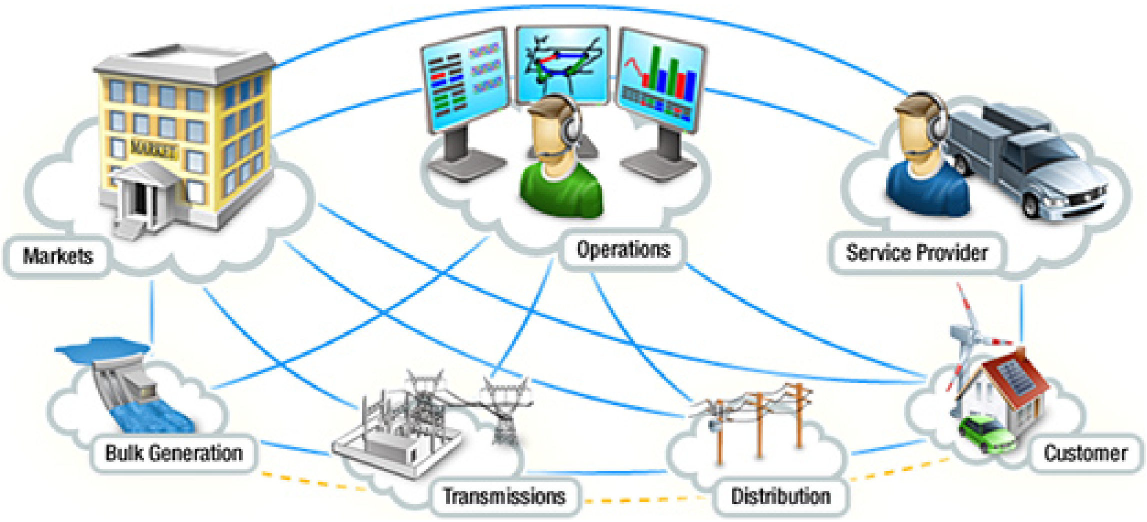


Figure 1.1 Schematic of IEEE Smart Grid conceptual model [12]

Smart Grid is still in its beginning stage, though it has started showing some impressive results, such as overall energy consumption reduction and peak load reduction. For example, Pacific Gas & Electric in US achieved 16.9% and 20% reduction in overall energy consumption through Smart Rate and energy efficiency program. Oklahoma Gas and Electric realised a 2% annual energy reduction by using voltage/volt-ampere reactive (Volt/VAR) control without any customer involvement [14].

This thesis concerns the impact of the Smart Grid in transforming the traditional electricity distribution network by adding a series of new technologies, such as new back-end IT systems, smart meters, smart sensors and communications network [14-19].

Global and national trends are beginning to affect the entire power sector. For example, the Australian Government is trying to reduce reliance on fossil fuels and strengthen investments in renewable energy sources because of the environmental issues. Ageing electric infrastructure will require costly upgrades to meet the demands of increasing customer load and expanding modern economy. It predicted that Australian electricity prices will double in five years [20]. The nation will need to manage power more efficiently and effectively, lower the ratio of electricity consumption per economic output, reduce overall greenhouse gas emissions and encourage energy efficiency, improve reliability, and reduce costs [21].

The global and national trends have initiated a wave of innovation in distributed power generation, electric transport, energy efficiency and Smart Grid capabilities. Power utilities and solution providers across Australia and nations around the world are starting to experiment and deploy a wide range of these innovations.

Demand side management targets the reduction of peak demand as well as conservation and efficient usage of energy. It helps to reduce expenditure on new network capacity and keep electricity prices at a reasonable level for consumers. The reduced overall energy consumption will lead to a decrease in greenhouse gas emission. Possible ways of demand side management include shifting energy usage to off-peak times, agreements to reduce demand at peak times upon request, substituting sources of generation, installing more efficient equipment, switching fuels and direct load control.

Electric vehicles will be an important load group serviced by the future Smart Grid. Renewable energy sources can be used to provide a large portion of vehicle energy consumption. Electric vehicles and pluggable hybrids can provide hundreds or even thousands of megawatt-hours controllable load and storage that could readily be matched to intermittent generation sources such as wind and solar. The transmission and distribution system could be made available for electric vehicle charging during off-peak periods, and electric vehicles can potentially be powered by renewable energy sources [22-25].

The nature of the distribution system is changing. Network designers are facing great challenges in this new situation. To bring the Smart Grid vision to reality, Australia will need to integrate information processing and communications into power systems to create a unified Smart Grid that includes generation, transmission, distribution, retail and end-use. Smart Grid encompasses a broad range of technologies and incorporates a broad range of new equipment supplied by various manufactures. It includes a suite of applications which are currently at different stages of technical and economic maturity [21, 26]. The selection of the correct technology will be situation dependent [27].

In Australia, Smart City – Smart Grid is a major initiative by the Federal Government that is testing the following Smart Grid capabilities.

1) Customer-side applications

- Information (e.g. energy usage or CO_2 information provided by website or in-home display)
- Controls (e.g. in-home displays, automated controls for appliances, programmable thermostats with communications)
- Tariffs that fluctuate with time of usage (e.g. Time of Use, Critical Peak Pricing, Real Time Pricing).

2) Key enabling applications

- Smart metering infrastructure (SMI) - known as advanced metering infrastructure (AMI) in Victoria.

3) Grid-side applications

- Integrated Volt-VAR control (including conservation voltage reduction)
- Fault detection, isolation and restoration (FDIR)
- Substation and feeder monitoring and diagnostics
- Wide-area measurement.

4) Renewable, distributed energy and electric vehicles

- Distributed storage (which may include electrical vehicle elements)
- DG enablement (e.g. solar photovoltaic (PV) on residential roof tops)
- Electric vehicle support.
- Data collection, processing and back-office

1.2 Statement of the problem

Distribution networks are growing with complexity, due to:

- Changes in demand profiles with increases in short term peak demands, especially in distribution feeders;
- Unexpected changes in energy requirements. In several jurisdictions, energy demand has reduced in the last two years;
- The emergence of bidirectional power flows and voltage rises due to the massive deployment of PV. Australia passed 1,000,000 roof top systems and 2400MWp in March 2013.
- New load groups including electric vehicles.

Operational changes and increasing complexity are a challenge to the current computational techniques and analysis tools. Existing deterministic modelling and simulation tools are well suited to the assessment or design of a single feeder, but do not automatically provide an insight into the implications of a broad scale technology deployment across an extensive network [27, 28]. To work in the Smart Grid environment, more sophisticated models and simulation tools are needed.

Australian distribution engineers commonly classify distribution feeders as CBD, Urban, Rural Short (< 200km) and Rural or Rural Long (> 200km), though it is not clear how these classifications have arisen. There has been no systematic statistical analysis of Australian distribution feeders. The distribution network impact assessment of changes in load behaviours or new technology deployments requires a rigorous understanding of the features of the network itself. Statistically representative feeder test cases and load models are required if the simulation results are to be reflective of the actual network performance outcomes. This thesis focuses on developing prototypical feeder models and domestic load models.

A formal method starts by selecting a small number of “prototypical feeders”, which can be statistically representative of the feeders encountered in a given network. There is an emerging demand for these models from researchers for assessing smart grid technologies. A compact but statistically representative test set allows distribution researchers and engineers to quickly evaluate smart grid technologies with confidence that the results will be meaningful in the physical network. Once a statistically significant set of feeders is determined for any given network, it is possible to use these to efficiently evaluate a range of smart grid technologies that could be applied. It becomes possible to match a set of technologies to specific feeder classes where the best technical and economic performance can be obtained.

There has been an international recognition of the need for representative test cases. MV test feeders have been provided by organisations, such as the Pacific Northwest National Laboratory (PNNL) of Department of Energy (DOE), Ausgrid, the United Kingdom Generic Distribution System (UKGDS) and IEEE [29-37]. The LV network is responsible for the larger part of the consumer voltage drops and phase imbalance. Few representative LV feeder studies have been found in the literature. This may be due to the less complete documentation of LV feeders and the sheer size of the data bases.

Australian engineers, software programmers and researchers need feeder models for the same reasons they are needed internationally to test technologies, especially important for MV level technologies, such as switched capacitors, integrated volt/VAR controller (IVVC), larger distributed generation (DG), dSTATCOMS. Overseas feeder MV models, especially IEEE test feeders, have been frequently used in Australia.

Some models are reasonable, but others, such as IEEE 13 bus system which used 4.16kV branches and components, are irrelevant to the Australian setting.

As for the LV models, there is a critical need. There are no published representative models. If models were developed there are specific geographical peculiarities. The US LV system is completely different from Australia. Models based on this system will provide no meaningful basis by which an Australian utility or researcher could evaluate LV impacts of EV or solar. Australia has three-phase four-wire system, while US consumers are supplied at 120-0-120V single phase or three phase arrangements which provide a 240V supply.

In the LV system, there are limited models for consumer load. Most utilities develop a “deemed profile” for residential customers and this is normally determined by the observation of group demands on distribution transformers or medium voltage feeders with a dominance of residential load. The LV networks are designed to provide adequate voltage regulation at the periods of peak demand and the design is often based on a single agreed demand figure for a residential consumer. These approaches cannot incorporate a more detailed understanding of consumer load across the diurnal cycle and have led to design approaches that contribute to power quality problems such as voltage rise in systems with high penetrations of PV. This thesis will seek to develop a much more nuanced understanding of consumer load and to present models that can support the analysis of new technologies.

1.3 Research objectives

This work is aimed towards an improved statistical understanding of distribution feeders, an improved probabilistic understanding of their loads and methods for network-wide assessment of specific technologies. The primary research contributions relate to fundamental knowledge for distribution system modeling as follows:

- Rigorous methods of establishing network test cases that, as a compact limited membership set, are representative of a particular large and complex distribution network;

- Improved understanding of the topological partitioning of Australian MV and LV distribution networks;
- New methods of feeder data selection, especially parameter selection, to ensure efficient and effective statistical clustering outcomes;
- Improved probabilistic daily load models with high PV penetration.

The secondary research objectives relate to informing distribution engineering design practice as follows:

- Efficient methods for establishing small numbers of test cases which characterise a specific network facilitating the quick assessment of network-wide benefits for specific technologies.
- The deployment of risk-based assessment methodologies for distribution network capacity and power quality including;
    - Methods to establish accurate design values for load diversity especially in respect to new demand side strategies and load groups;
    - Methods to establish design values for the firm capacity for intermittent sources in distribution networks.

1.4 Significance

This thesis makes contributions to knowledge under the two main areas - representative feeder models for MV/LV distribution networks and load models for residential households with high PV penetration.

A rigorous assessment of any new technology first requires the establishment of test cases. Studies have been published on the application of cluster analysis to MV feeders to extract representative or prototypical feeders, which has an excessive input data requirement in the PNNL case. This thesis developed a multivariate statistical method of higher utility by using fewer information rich variables that can be easily extracted from common DNSP data bases. For the LV feeder test cases, there is no published literature that comprehensively addresses statistical studies on LV feeders. The thesis developed parameter selection approaches to assist in extracting representative feeder subsets from large LV feeder sets.

This thesis developed load models for residents with high PV penetration, which have not been statistically studied in the literature. The reconstructed daily load profiles, PDFs and CPDFs fits well with the empirical load data.

1.5 Thesis structure

Chapter 2 gives introduction on network characterisation and to the Perth Solar Cities Program that gave an ideal opportunity to collect fine scale consumer and solar data and to examine the correlation of generation and load.

Chapter 3 presented a highly efficient multivariable statistical analysis method. It relies upon a few key variables which are highly meaningful from an engineering perspective and readily available in most if not all distribution companies. This chapter firstly introduced the statistical cluster analysis used in the distribution feeder classification, the main purpose of which is to find out the optimal number of clusters to be used. Discriminant analysis was then conducted based upon the result from the cluster analysis. Combining the two analyses makes it a stable multivariable statistical analysis method, which attempts to obtain a stable classification results with minimum error rate. A set of discriminant functions can be extracted from the method as feeder classifiers.

Chapter 4 applied the methodology in the previous chapter for the representative MV and LV feeder studies. The designs should be economic rather than too conservative. Improved tools needed to determine ratings and equipment with uncertain loads. This chapter presents case studies to identify statistically significant feeders with a limited number of engineering variables using two very different data bases – a MV feeder data base of 204 members and an LV data base of 8,858 members. Distribution utilities maintain data bases that can provide relevant clustering information. The capability of the general method to form useful cluster classifications in such disparate applications illustrates the universality of the method.

Chapter 5 presented a method to develop a hybrid model for high demand days modelling of consumer load. A model for underlying diurnal variation was constructed using a truncated Fourier series representation. Cluster and discriminant analysis were used to extract six representative daily load profiles. After the low frequency variation was established a high frequency residual remains. This was modelled using a power spectrum. The load models were then validated using the data from the five high demand test days. For each test day load ensembles were constructed that included appropriate numbers of residential loads of each cluster type. The load groups successfully produced aggregated load behaviours that compare well with the physical load data in terms of load profile, PDFs and CPDFs.

In Chapter 6, in comparison to the high demand day models, load models were built and tested when the aggregated loads recorded from the transformer side were the lowest. The load profiles, PDFs and CPDFs were found to compare well with the physical load data.

Finally, Chapter 7 summarises the work in the thesis and presents discussions about possible directions in future research.

References:

[1] S. M. Amin, and B. F. Wollenberg, 'Toward a smart grid: power delivery for the 21st century', Power and Energy Magazine, IEEE, vol. 3, no. 5, pp. 34-41, 2005.

[2] X. Mamo, S. Mallet, T. Coste, and S. Grenard, 'Distribution automation: The cornerstone for smart grid development strategy', in Power & Energy Society General Meeting PES '09, IEEE, 2009, pp. 1-6.

[3] J. Heckel, 'Smart substation and feeder automation for a SMART distribution grid', in 20th International Conference and Exhibition on Electricity Distribution, 2009, pp. 1-4.

[4] W. Xu, Y.-h. Zhou, and J.-l. Zhu, 'Energy-efficient distribution in smart grid', in International Conference on Sustainable Power Generation and Supply SUPERGEN '09, 2009, pp. 1-6.

[5] H. Farhangi, 'The path of the smart grid', Power and Energy Magazine, IEEE, vol. 8, no. 1, pp. 18-28.

[6] F. Jiyuan, and S. Borlase, 'The evolution of distribution', Power and Energy Magazine, IEEE, vol. 7, no. 2, pp. 63-68, 2009.

[7] K. I. Geisler, 'A smarter greener power grid', in Power Systems Conference PSC '09, IEEE, 2009, pp. 1-3.

[8] A. Ipakchi, and F. Albuyeh, 'Grid of the future', Power and Energy Magazine, IEEE, vol. 7, no. 2, pp. 52-62, 2009.

[9] M. Paun, and G. Lorenz, 'Smart grids and networks of the future - Eurelectric views', in 20th International Conference and Exhibition on Electricity Distribution, 2009, pp. 1-4.

[10] S. E. Collier, 'Ten steps to a smarter grid', in Rural Electric Power Conference REPC '09, IEEE, 2009, pp. B2-B2-7.

[11] 'Grid 2030 – A National Vision for Electricity's Second 100 Years', U.S. Department of Energy, Office of Electric Transmission and Distribution.

[12] IEEE. 'Smart Grid Conceptual Model ', 2013; http://smartgrid.ieee.org/ieee-smart-grid/smart-grid-conceptual-model, accessed Feb 6th 2013.

[13] 'SmartGrids SRA 2035 Strategic Research Agenda Update of the SmartGrids SRA 2007 for the needs by the year 2035'; www.smartgrids.eu/documents/sra2035.pdf, accessed Feb 6th 2013.

[14] GridWise Alliance Project Team and Quanta Technology Project Team, 'Realizing the value of an optimised electric grid', http://www.gridwise.org.

[15] 'Strategic Research Agenda for Europe's Electricity Networks of the Future'; http://www.smartgrids.eu/documents/sra/sra_finalversion.pdf, accessed Feb 6th 2013.

[16] 'Towards Smart Power Networks – Lessons learned from European research FP5 projects,' http://ec.europa.eu/research/energy/pdf/towards_smartpower_en.pdf, 2005, accessed Feb 6th 2013.

[17] S. Bruno, S. Lamonaca, M. La Scala, G. Rotondo, and U. Stecchi, 'Load control through smart-metering on distribution networks', in PowerTech, IEEE Bucharest, 2009, pp. 1-8.

[18] J. Deuse, K. Purchala, and G. Bourgain, 'Distribution network design, management and regulation', in SmartGrids for Distribution, IET-CIRED, CIRED Seminar, 2008, pp. 1-4.

[19] M. LeMay, R. Nelli, G. Gross, and C. A. Gunter, 'An Integrated Architecture for Demand Response Communications and Control', in Proceedings of the 41st Annual International Conference on System Sciences, Hawaii, 2008, pp. 174-174.

[20] A. Kohler. 'Australia faces double shock on electricity prices'; http://www.smartcompany.com.au/climate-change/20091109-australia-faces-double-shock-on-electricity-prices-kohler.html, accessed Feb 6th 2013.

[21] 'Smart Grid Smart City - A new direction for a new energy era'; http://www.environment.gov.au/smartgrid/pre-deployment.html, accessed Feb 6th 2013.

[22] U. K. Madawala, P. Schweizer, and V. V. Haerri, '"Living and mobility" - a novel multipurpose in-house grid interface with plug in hybrid BlueAngle', in IEEE International Conference on Sustainable Energy Technologies ICSET, 2008, pp. 531-536.

[23] M. Ferdowsi, 'Vehicle fleet as a distributed energy storage system for the power grid', in Power & Energy Society General Meeting PES '09, IEEE, 2009, pp. 1-2.

[24] I. Cvetkovic, T. Thacker, D. Dong, G. Francis, V. Podosinov, D. Boroyevich, F. Wang, R. Burgos, G. Skutt, and J. Lesko, 'Future home uninterruptible renewable energy system with vehicle-to-grid technology', in Energy Conversion Congress and Exposition ECCE, IEEE, 2009, pp. 2675-2681.

[25] S. G. Wirasingha, N. Schofield, and A. Emadi, 'Plug-in hybrid electric vehicle developments in the US: Trends, barriers, and economic feasibility', in Vehicle Power and Propulsion Conference VPPC '08, IEEE, 2008, pp. 1-8.

[26] K. P. Schneider, D. Chassin, Y. Chen, and J. C. Fuller, 'Distribution power flow for smart grid technologies', in Power Systems Conference and Exposition PSCE '09, IEEE, 2009, pp. 1-7.

[27] B. Saint, 'Rural distribution system planning using Smart Grid Technologies', in Rural Electric Power Conference REPC '09, IEEE, 2009, pp. B3-B3-8.

[28] P. Wolfs, and S. Isalm, 'Potential barriers to smart grid technology in Australia', in Australasian Universities Power Engineering Conference AUPEC'09, 2009, pp. 1-6.

[29] R. Singh, B. C. Pal, and R. B. Vinter, 'Measurement Placement in Distribution System State Estimation', Power Systems, IEEE Transactions on, vol. 24, no. 2, pp. 668-675, 2009.

[30] P. Punjad, G. W. Ault, and J. R. McDONALD, 'Impacts and management arrangements for high penetration distribution generation', in 19th International Conference on Electricity Distribution (CIRED), Vienna, 2007, pp. 21-24.

[31] E. M. Davidson, M. J. Dolan, G. W. Ault, and S. D. J. McArthur, 'AuRA-NMS: An autonomous regional active network management system for EDF energy and SP energy networks', in Power and Energy Society General Meeting, 2010 IEEE, 2010, pp. 1-6.

[32] K. P. Schneider, Y. Chen, D. Engle, and D. Chassin, 'A taxonomy of North American radial distribution feeders', in Power & Energy Society General Meeting PES '09, IEEE, Power & Energy Society General Meeting, 2009. PES '09. IEEE, 2009, pp. 1-6.

[33] Y. Li, and P. Wolfs, 'A statistical study on topological features of high voltage distribution networks in Western Australia', in Universities Power Engineering Conference (AUPEC), 20th Australasian, 2010, pp. 1-6.

[34] K. P. Schneider, J. C. Fuller, F. K. Tuffner, and R. Singh, Evaluation of Conservation Voltage Reduction (CVR) on a National Level, Pacific Northwest National Laboratory report, July 2010.

[35] W. H. Kersting, 'Radial distribution test feeders', Power Systems, IEEE Transactions on, vol. 6, no. 3, pp. 975-985, 1991.

[36] W. H. Kersting, 'Radial distribution test feeders', in Power Engineering Society Winter Meeting, IEEE, vol.2, 2001, pp. 908-912.

[37] Ausgrid. 'Smart Grid, Smart City'; http://www.smartgridsmartcity.com.au/, accessed Feb 6th 2013.

Chapter 2 Distribution network and domestic load characterisation for a Smart Grid environment

2.1 Introduction

The Australian power industry is very large and complex, with $11 billion in revenue, over 45,000 kilometres of transmission lines and 700,000 kilometres of distribution networks, and over nine million customers, including many in remote areas [1, 2]. A distribution feeder has typically 30 to 100 distribution transformers. The residential distribution feeder typically operates at 11kV or 22kV and services up to a few thousand households. The network could be characterised at three levels. The framework of the work in this thesis is shown in Figure 2.1. The key steps are modeling developments for:

- The MV feeder level – A network typically contains a few hundred MV feeders that are relatively well documented. Representative feeders can be selected to provide relevant test cases for use with a range of analysis tools.
- The LV feeder level – There are over ten thousands of LV feeders in a typical distribution network. Representative test cases again need to be extracted to provide and input into a range of analysis tools
- Diurnal load models – In many smart grid applications, a nuanced understanding of the intra-day load variation is needed for domestic consumers. These models can be constructed from domestic load data sets. If the specific impacts of distributed generation are to be determined, then that data is required as well.

Figure 2.1 illustrates how these elements might be combined and used to construct a Monte Carlo simulation. This is one example, amongst many, that would be appropriate if estimations were to be made of the statistical variability of network operation as a means to estimate power quality performance.

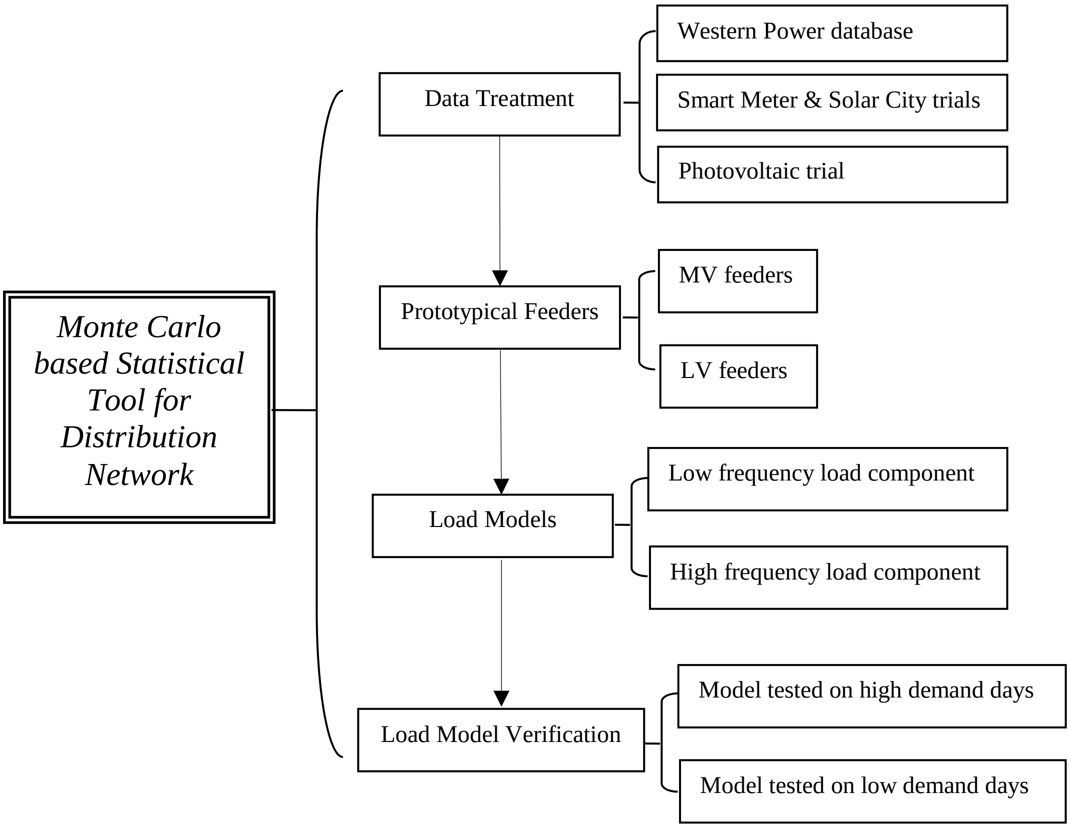


Figure 2.1 A scheme showing the framework of research

2.2 Literature review

2.2.1 Distribution feeders test sets

Representative test sets for MV feeders have emerged in the UK, the USA, and Australia [3-11]. The UKGDS provides a range of resources for the simulation and analysis of the impact of distributed generation [3]. UKGDS provides test feeders that allow the consistent comparison of novel technical solutions. In this case, a consultation process with experienced network personnel was used to identify representative systems.

PNNL of DOE conducted a study of US feeders to develop a set of 24 prototypical feeders using the method of hierarchical clustering [6, 12]. PNNL has recently published an assessment of conservation voltage reduction benefits based on their North American feeder taxonomy [8]. This recent example applied the PNNL test feeder set to provide a rapid, and statistically defensible, evaluation of Conservation Voltage Reduction (CVR), across the US distribution network.

IEEE test sets were presented and quoted very often. It published the complete data for three four-wire wye and one three-wire delta radial distribution test feeders, which could be used by program developers and users to verify the correctness of their solutions [10]. An updated version of the IEEE same test feeders along with a simple system that can be used to test three-phase transformer models were published then. It presented the complete data for three four-wire wye, one three-wire delta test feeders and a simple feeder for testing three-phase transformer models [11]. The data could be used to test software using these test feeders.

In this thesis, chapter 3 and chapter 4 aim to build representative MV/LV feeder models in order to identify statistically significant feeders with a limited number of engineering variables.

2.2.2 Residential load models and profiles

The residential households consume substantial amount of energy. In power systems, Residential electricity consumption takes up to 16% to 50% of all the electricity consumption, and averages approximately 30% worldwide [13]. Yet, it is less well understood compared to other sectors such as commercial, industrial, agriculture and transportation.

Modelling of residential loads becomes complex when taking account into the stochastic nature of load. It varies with time and customer wealth. It is strongly affected by the weather conditions and these drive heating and cooling demands. [14-16]. The addition of PV generation in the system introduces further daily and seasonal changes that make the residential loads more difficult to model accurately. The vast majority of load modelling has been directed to the modelling or residential consumer groups. In this thesis it will be shown that individual consumer loads fluctuate wildly while the aggregate load of groups of a few dozen consumers is much less volatile. The modelling of individual customers is challenging but necessary for certain classes of analysis, such as the deployment of residential scale storage.

The modelling methods for residential electricity consumption can be summarised into two main groups - ‘top-down’ and ‘bottom-up’ methods [13] .The ‘top-down’ approaches utilise historic aggregate energy values and regresses the electrical usage as a function of variables such as macroeconomic indicators and energy price, etc..

The widely used ‘top-down’ models are econometric and technological models [17-24]. Some of the models utilise a combination of the econometric and technological models. As to the econometric models, they are mainly based on characteristics as price and income. Technological models analyse broader characteristics, such as climatic conditions, appliance ownership and number of customers.

As the top-down models only need aggregate data, the advantages are simplicity and the data are widely available. As these methods do not involve modeling on individual households and cannot reflect on the impact by new technologies, so they can be used for supply analysis based on long-term projections of electricity demands.

On the other hand, the ‘bottom-up’ approaches refine the understanding of the details on household load. These methods calculate the residential energy consumption of individual or groups of houses and then extrapolate these results to represent the region or nation [25-30]. The ‘bottom-up’ models include statistical and engineering models.

The statistical methods analyse characteristics of some sample households, such as macroeconomic, energy price and income, to build relationships between the sample and region. The wide-published statistical techniques are regression, conditional demand analysis (CDA) and neural networks.

Engineering methods calculate energy consumption by using parameters of power ratings and/or heat transfer and thermodynamic principles. These methods involve developing a representative housing database estimating the energy consumption of the houses using a building energy simulation program. Database representative of the housing stock with detailed house description data are needed, as well as extensive user expertise and lengthy input data preparation time. However, just because of this, they can be used to evaluate the impact of a wide range of technologies.

A big advantage of the ‘bottom-up’ approach is that it explicitly identifies occupant behaviour and household renewable energy generation such as solar panels. However, compared to the ‘top-down’ approaches, the ‘bottom-up’ approaches can be complex due to greater data sets needed.

Some of the above models focus on peak demand while some seek to develop load profiles. Demand models that focus on the maximum demand estimation cannot reflect the diurnal load variation due to PV generation.

The work in this thesis applies a top-down approach based on smart meter data sets from the Perth Solar City high penetration PV trial. To build accurate residential load profile models, it is important to take into account of the following variability of individual loads that can be significant and rapid [31], and the thesis aims to incorporate these factors in the load models.

- Diurnal variations;
- Intra-daily variations;
- Variations between different customers;
- Seasonal variations.

The widely researched electrical load profiling techniques are Gaussian and mixed Gausian process models, Fourier transform based models, neural networks, conditional demand analysis (CDA), fuzzy logic, autoregressive models, and wavelet based models . Each method has its each own advantages and disadvantages.

A common profiling technique is based on Gaussian process [32-35]. A Gaussian process is a generalization of the Gaussian probability distribution. It is a collection of random variables, which have a joint multivariate Gaussian distribution [33]. It can be determined with relatively small training points if used for forecasting, and provides a simple representation of the load profile, which can be expressed only by the covariance and mean. When facing a nonstandard distribution, Gaussian mixture model can be used through a weighted combination of normal distributions [34]. In this case the expectation maximization algorithm can be applied to obtain the parameters of the Gaussian mixture. The algorithm finds out the maximum-likelihood estimate of the parameters of a distribution when the data are incomplete. The Gaussian mixture method was tested on a UKGDS generic distribution network model, which comprises 95 buses with 55 load buses and two wind farms as sources of distributed generation [34]. The load probability distribution was generated by segmenting the range of the data into various disjoint categories for each load bus. Another advantage of Gaussian process is that it deals spiky load profile with high accuracy. Also, computational tools can easily incorporate the Gaussian distribution. But it is less good at handling smoother load profiles.

Fourier transform is a powerful tool for the numerical parameterisation of load profiles by calculating Fourier series approximations to the profiles so that the coefficients of series can represent the shapes. The Fourier transform was found to have better approximation in load profiling compared to polynomial-type sets of orthogonal functions, such as Bessel functions, Legendre functions, associated Legendre functions, Chebyshev polynomials of types I and II [36].

The main advantage of Fourier transform techniques is that it is able to model both the temporal and magnitude components of load [36-38]. The coefficients of series can not only represent the load profile shapes, but also have some physical meanings. The constant term is proportional to the area under the load profile that is the total power consumed; and the relative values indicate some profiling information. For example, if the coefficient of the leading cosine term dominates, the load shape will be cosine-like, which would indicate high power consumption at night.

The Fourier transform techniques is most commonly applied using the discrete Fourier Transform, (DFT), is usually carried out by the algorithm of Fast Fourier Transform, (FFT). Spectral leakage can be a problem in the technique that could be fixed by ‘windowing’- it multiplies the time series data with a window whose transform has reduced side lobes [37]. To decompose profile into frequency spectra, a technique called ‘super-resolution’ could be used. An improved super-resolution method called modified Papoulis super-resolution was found to be good at locating true components within the noise and has shown better performances in term of spectral identification and extrapolation. The disadvantage of the Fourier transform technique is its accuracy drops when characterising spiky load profiles [39].

Neural network techniques are simplified mathematical models of biological neural networks. The term was traditionally used to refer to a network or circuit of biological neurons. The technique allows all end-uses to affect one another through a series of parallel neurons. As it is a parallel model, unlike the Fourier transform, the coefficients have no physical significance [13].

At first, neural network techniques were mainly used for utility load forecasting and to predict energy consumption of individual buildings [40, 41]. Then the Joint Center for Energy Management at the University of Colorado started applying them for energy analysis in 1998 [42]. Issa also introduced the application of neural network modelling to the residential sector [43]. This limited use possibly due to neural network’s data requirements and the lack of physical meaning of the coefficients. After that, they have been used to predict energy consumption of individual households, as they are good at identifying the behaviour of non-linear systems [44-50]. The approaches can give accurate modeling on the space heating energy consumption, and the domestic hot-water heating energy consumption [51], as well as lighting, and space-cooling consumptions. They are also able to model on national or regional residential energy consumption [52].

Neural network has shown very good capability in evaluating the actual consumption model compared to some other techniques. For instance, Sh. MohammadZadeh compared neural networks and econometrics models [53]; and M. Aydinalp compared them with engineering models [54]. The approaches could be used as a substitute for statistical approaches due to their simplicity and accuracy. However, the black box approaches makes them difficult to establish a relationship between the input and output [55].

CDA is a regression-based method that performs regression based on the presence of end-use appliances. It does not require detailed data - an appliance survey from the occupant and energy billing data would be sufficient. It exploits the differences in ownership to determine each appliance's component of the total consumption. However, since these are regression-based models, the number of households needs to be large, so dataset with a variety of appliance ownership is needed. Data from hundreds or even thousands of households are required in order to obtain reliable results depending on the number of variables used. However, the models do not provide much detail and flexibility. Also, multi-collinearity problems often make it difficult to isolate the energy use of highly saturated appliances, such as the main refrigerator (saturation approximately 100%). The size and number of refrigerator affect the energy consumption.

In fuzzy logic techniques, the relationship between input and output is clearly defined [55-57], but the number of variables required to characterise the output is usually large, and the temporal component is not characterised unless each half hour period is characterised separately.

Wavelets techniques separate the electricity load profile into high and low frequency components before applying the wavelets transform. The techniques allow each frequency component to be considered with an appropriate temporal resolution. The discrete wavelet transform is good at handling non-stationary discrete signals in both time and frequency domains. For the disadvantage of Wavelets, the number of variables involved in the model is doubled after the separation into two components [58-61].

In autoregressive techniques, the variable coefficients vary significantly with small changes in load profiles, which make it difficult to group or compare customers [36]. As a common autoregressive technique, Markov chains can characterise the high variability of residential load. For example, O. Ardakanian built models for households during on-peak periods, off-peak periods, and mid-peak periods based on fine-grained (i.e. every six seconds) measurements of energy consumption in 20 homes over four months [62]. However, they requires characterising at least each half hourly, leading to a large number of variables [63].

Regression models regress the aggregate residential energy consumption onto parameters which are expected to affect energy consumption. Input variables that have little effect can be removed for simplicity. Based on the combinations of inputs, the model's coefficients may or may not have physical significance [13].

Within the probabilistic analysis framework, the variability of the electricity consumption of an individual house generally depends on the presence at home of family members and on the time of high-power appliances used with relatively short usage duration. Researches have been done in identifying suitable probability distributions among beta, exponential, gamma, Gumbel, log-normal, normal, Rayleigh and Weibull. The gamma, log-normal and normal probability distributions have found to be the most suitable ones after goodness-of-fit tests [64].

Using the probabilistic analysis, Harris studied the dynamic relationships between electricity consumption and several potentially relevant variables, such as weather, price, and consumer income finding a high seasonality of electricity demand [65]. Ranjan developed multiple linear regression models of energy consumption for different seasons [66]. Amarawickrama presented a time series analysis of electricity demand in Sri Lanka [67]. They studied different time series estimation methods in terms of modelling past electricity demand, estimating the key income and price elasticities. After comparing six econometric techniques, the forecasts generated from the six models do not differ significantly up until 2025, therefore, the Sri Lanka electricity authorities could have some faith in econometrically estimated models. Cagni reported a bottom-up approach using probabilistic technique to combine the load consumption of the aggregated customers in Italy [68]. The two-phase approach evaluated the aggregated load diagrams of a set of single house customers in extra-urban areas. The first phase of the approach made a direct investigation into the customers' electricity consumption, which was related to the presence at home of family members and to the periods of usage of the appliances. In the second phase, an overall simulation was carried out to assess the time-dependent variation bands of load power in function of the number of customers. The simulations have been performed by running 100 cases for each number of customers.

The advantage of multiple regression technique is that it is ideal for generating variable load, but load profiles with high accuracy need to be characterised separately at least half-hourly [31, 64, 69]. Monte Carlo simulation technique is one of the most commonly used probabilistic methods. More details on this will be given in the next section.

Last but not least, there are hybrid approaches that combine two or more techniques, for instance, a combination of Fourier transform with neural network [38], where Fourier transform described the periodic fluctuation and neural network predicted the monthly trend. Least squares method was used to adjust on the number of data used in the Fourier transform. This independent prediction of fluctuations and trend ought to provide a more successful forecasting than the use of single technique, and the hybrid model has shown more accurate prediction than only using neural network. M. Aydinalp-Koksal proposed it may be possible to develop a hybrid model that uses the engineering model for physical and thermodynamic modelling whereas neural network models for modelling of socio-economic factors [51]. There are publications on a combination of wavelet decomposition and neural networks that used wavelet decomposition and neural networks to capture the features of load at low and high frequencies [59, 70]. To address the shortcomings of both the engineering and the statistical based models, L. Swan has developed a hybrid engineering model and neural model for the Canadian housing sector [71]. The neural model was applied for the highly occupant sensitive domestic hot water, appliances and lighting energy consumption, while engineering model was used to predict the space cooling and space heating energy consumption. Overall, all the hybrid approaches have shown promising modelling results.

This thesis built a hybrid load model based on Fourier transform and probabilistic technique. The model also applied the statistical cluster and discriminant analysis into the load profiling process in order to select Fourier series coefficients. Monte Carlo simulation was utilised in the model, as it is good at producing variable residential load profiles. As it mentioned before, it requires greater data sets at least at half hour intervals. The thesis made use of data from the PV trial, which is a part of the Perth Solar City program. During the trial, consumer smart meter data have been made available on 15 minute intervals. This work seeks to identify specific load types within a set of residential load data to provide an improved understanding of the load structure. This is lost in the stochastic approaches. Modeling accuracy depends on the load group and the model. The use of the overlap of the Smart Meter deployment and the high penetration PV trial examined the agreement of statistical load modeling and identified the impact on loads when with and without PV generation. It also allowed studies to look into the correlation of generation and load.

2.2.3 Probabilistic approach to distribution network design

Distribution design is often based on "After Diversity Maximum Demand (ADMD)" concepts. It estimates the maximum likely voltage drop accounting for consumer diversity. When a new transformer added to a network without the knowledge of the loading, the common method uses an empirical approach given by Boggis [72]. It is derived from the recorded maximum yearly nodal demand on the transmission network divided by the number of customers serviced. The ADMD is defined as the maximum coincident total demand of a homogeneous group of consumers divided by the size of the group:

$$ADMD = \frac{MD}{N} \tag{2.1}$$

However, diversity in customers' loads makes the maximum demand (*md*) of a group of *N* similar loads smaller than the total of their individual maximum demands. The ADMD is multiplied by a diversity factor (*DF*) that increases the demand per customer as the size of the group decreases.

$$ADMD \times DF = \frac{MD}{N} \tag{2.2}$$

As the number of customer serviced in a network increases, the *DF* steadies at a value. The value of *DF* reaches 1 when the group size approaches infinity.

$$DF = 1 + \frac{k}{N} \tag{2.3}$$

$$ADMD = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} P_n \tag{2.4}$$

where, P_n is the demand of customer at the time of system maximum demand; k is a constant that needed to be given empirically. With the development of traditional methodology, the formulae of ADMD and DF have been adapted to include more factors, such as consumers' type, month and day of the week.

It can be seen that the deterministic approach for the network calculations made use of suitable coincidence factors to combine the MV loads connected to each line. The coincidence factors depended mainly on the number of customers, they could lead to uneconomical or unreliable solutions, as the deterministic algorithms adopt the worst case approaches. ADMD calculates the likely maximum demand for a group of customers within a specific period of time, the time duration of the received voltage is not considered [73], and it does not take into account the stochastic nature of the demand.

Distribution system behaviour is stochastic process and varies with time. It can be affected by a large variety of factors. For example, temperature change has very significant impact to the power consumption of distribution feeders in the commercial and residential areas. C. S. Chen studied the effect of temperature change to the Taiwan power system load demand by using the typical load patterns of customer classes [74]. The power consumption of Taipei city increased by 366 MW or 22% when the temperature rises by 5 ºC, mainly introduced by air conditioners.

In this increasingly energy-conscience world, detailed and accurate characterisations of loads become important aiming at promotion of conservation, efficiency and new technologies. The accurate evaluation of such systems should be based on probabilistic techniques that respond to the behaviours. The distribution systems are affected by many uncertainties. For example, with increasingly new distributed generation installed the level of uncertainties that characterises the planning environment increases. A high degree of uncertainty introduced by renewable energy source requires probabilistic methods in distribution network calculations. Risks have to be explicitly considered in planning, such as power flow and network losses. Decision taking account of uncertain scenario can reduce power losses cost and defer utility investment.

There has been growing of probabilistic simulations in MV [75-88] and LV [14, 15, 73, 89-102] distribution networks. The applying of Monte Carlo simulations have shown that deterministic models of distributed generation are far too pessimistic in discounting intermittent sources and fail to fully value the contribution to power quality improvement. The impact of new technologies, such as an uptake of electric vehicles, roof top solar generation or an increased use of demand management is increasingly studied using probability based tools including Monte Carlo based methods. In this thesis, chapter 5 and chapter 6 aim at providing accurate load models by means of Monte Carlo simulations.

2.3 Data collection

Feeder clustering research requires access to feeder data sets. This thesis focuses on methods of data selection and pre-treatment that reduces the costs and time required to assemble the data and improve the clustering outcomes. The previously published PNNL MV feeder clustering approach gathered 37 feeder variables [6]. Some variables are obscure topological parameters that are difficult to assemble.

This thesis established that high quality results can be achieved with far fewer variables if these are carefully selected for their high information content. The key feeder variables relate to:

- Feeder voltage and capacity;
- Circuit length;
- Load variability;
- Voltage regulation;
- Customer numbers and customer installed capacity.

All data collection processes have been developed to align with data sets that are commonly held by distribution network operators. These data sets include:

- Construction data sets;
- SCADA system load recordings;
- System reliability information generated by maintenance systems.

This thesis drew test case data from the South West Interconnected System (SWIS). Large scale data sets obtained during the Smart Meter and Solar City trials. The work made use of the data that generated by the roll out of 8,000 smart meters in the SWIS. These data were complemented as necessary by consumer data collected at finer time scales to develop a statistical understanding of various consumer and loads.

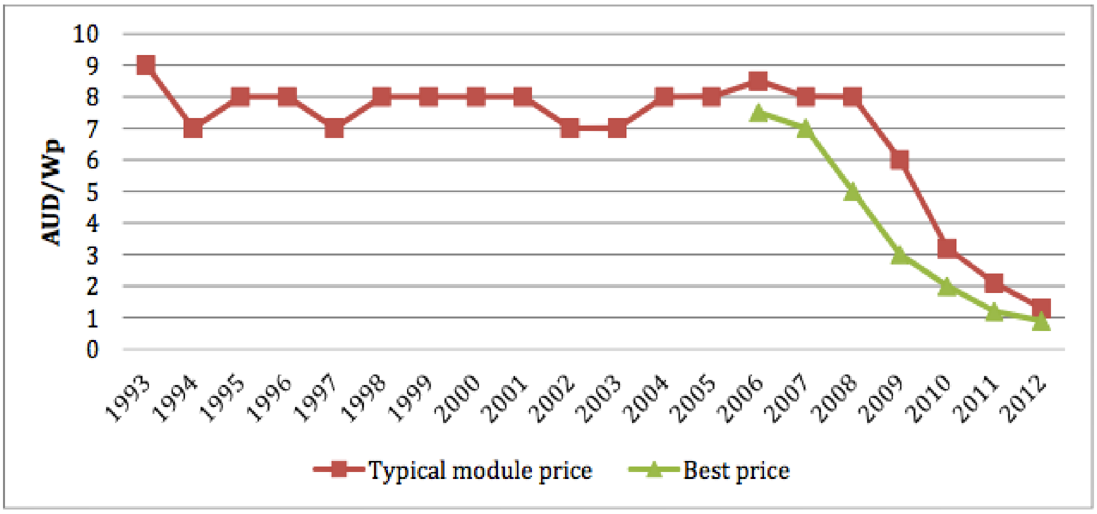
Western Power Corporation simultaneously conducted a high penetration PV trial under the Solar Cities Program. The project concentrated the available funding to secure a 30% to 40% penetration of domestic roof top solar systems on a specific feeder. This is an ideal opportunity to collect fine scale solar data and to examine the correlation of generation and load.

2.4 Solar Cities program

Australia's Solar Cities program is a $7 million innovative to promote solar power, smart meters, energy conservation and new approaches to electricity pricing in urban areas throughout Australia for a sustainable energy future [103, 104]. The Solar Cities are Adelaide, Alice Springs, Blacktown, Central Victoria, Moreland, Perth and Townsville. It is a partnership approach that involves all levels of Government, the private sector and the local community.

Perth Solar City Program was launched in Perth's eastern region in November 2009, and was managed by Western Power on behalf of the Australian Government. Perth Solar City engages all levels of government, businesses, not-for-profit organisations and community groups to address demand peaks and bring sustainable energy solutions to the Perth area. Large-scale PV systems have been installed at Perth Zoo, the Central Institute of Technology and Midland Foundry, and will benefit future projects in WA and elsewhere.

Perth Solar City program has trialled a number of energy efficiency initiatives designed to help residents reduce their energy use, save money and help the environment. Over 16,000 households have participated in the program, making it the WA's most comprehensive energy efficiency program. In total, 3,515 households received a home eco-consultation; 6,300 households received 12 months of eco-coaching; 700 homes were fitted with a SunPower PV system, and 1,100 homes purchased a Solahart solar hot water system [105].



Australian PV module prices 1993-2011 (current AUD)

Figure 2.2 Australian PV module prices from 1993 to 2011 in AUD
- courtesy of Arnold Mckinley ANU (Source: APVA 2012)

The Australian PV module price has been plummeting in the recent years as shown in Figure 2.2, and there has been large increase in the roof top PV installation in Australia, with the roof top installation of 2,450MWp in March 2013 as compared to 23MWp in 2008. PV is estimated to have produced 1,200GWh or 0.6% of the national requirement in 2011. The AEMO has forecast capacities are 5,100MW in 2020 and 12,000MWp in 2031 as shown in Figure 2.3.

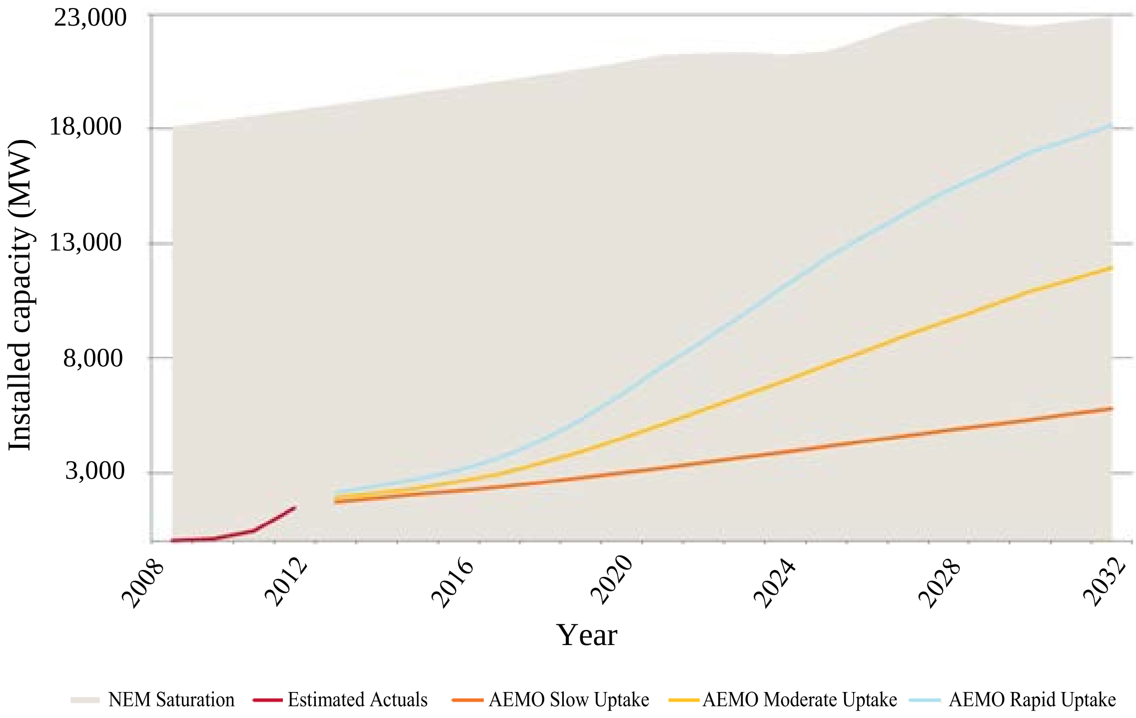


Figure 2.3 Roof top PV installed capacity forecasts for the NEM

Western Power, a West Australian TNSP/DNSP, conducted the high PV penetration trial with funding from the Australian Federal Government Solar Cities Program. The trial is located in a residential area, Pavetta Street, Forrestfield, Perth and focuses on a single feeder supplying 77 customers. This is a 230V/400V three phase four-wire LV feeder supplied from a 200kVA Dyn transformer. A total of 54kWp of roof top PV was installed by 34 customers. The trial provided an opportunity to collect consumer data in a high solar penetration environment for cluster analysis and load modeling.

References:

[1] 'State of the energy market 2008'; http://www.accc.gov.au/content/index.phtml/itemId/850040, accessed March 10th 2013.

[2] 'Smart Grid Smart City - A new direction for a new energy era'; http://www.environment.gov.au/smartgrid/pre-deployment.html, accessed March 10th 2013.

[3] R. Singh, B. C. Pal, and R. B. Vinter, 'Measurement Placement in Distribution System State Estimation', *Power Systems, IEEE Transactions on*, vol. 24, no. 2, pp. 668-675, 2009.

[4] P. Punjad, G. W. Ault, and J. R. McDONALD, 'Impacts and management arrangements for high penetration distribution generation', in 19th International Conference on Electricity Distribution (CIRED), Vienna, 2007, pp. 21-24.

[5] E. M. Davidson, M. J. Dolan, G. W. Ault, and S. D. J. McArthur, 'AuRA-NMS: An autonomous regional active network management system for EDF energy and SP energy networks', in Power and Energy Society General Meeting, 2010 IEEE, 2010, pp. 1-6.

[6] K. P. Schneider, Y. Chen, D. Engle, and D. Chassin, 'A taxonomy of North American radial distribution feeders', in Power & Energy Society General Meeting PES '09, IEEE, *Power & Energy Society General Meeting, 2009. PES '09. IEEE*, 2009, pp. 1-6.

[7] Y. Li, and P. Wolfs, 'A statistical study on topological features of high voltage distribution networks in Western Australia', in Universities Power Engineering Conference (AUPEC), 20th Australasian, 2010, pp. 1-6.

[8] K. P. Schneider, J. C. Fuller, F. K. Tuffner, and R. Singh, *Evaluation of Conservation Voltage Reduction (CVR) on a National Level*, Pacific Northwest National Laboratory report, July 2010.

[9] Ausgrid. 'Smart Grid, Smart City'; http://www.smartgridsmartcity.com.au/, accessed March 10th 2013.

[10] W. H. Kersting, 'Radial distribution test feeders', *Power Systems, IEEE Transactions on*, vol. 6, no. 3, pp. 975-985, 1991.

[11] W. H. Kersting, 'Radial distribution test feeders', in Power Engineering Society Winter Meeting, 2001. IEEE, 2001, pp. 908-912 vol.2.

[12] K. P. Schneider, Y. Chen, D. P. Chassin, R. G. Pratt, D. W. Engel, and S. E. Thompson, *Modern grid initiative-distribution taxonomy final report*, Pacific Northwest National Laboratory, 2008.

[13] L. G. Swan, and V. I. Ugursal, 'Modeling of end-use energy consumption in the residential sector: A review of modeling techniques', *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 1819-1835, 2009.

[14] R. Herman, and C. T. Gaunt, 'A Practical Probabilistic Design Procedure for LV Residential Distribution Systems', *IEEE Transactions on Power Delivery,* vol. 23, no. 4, pp. 2247-2254, 2008.

[15] J. K. R Herman, 'The statistical description of grouped domestic electrical load currents', *Electric Power Systems Research,* vol. 27, pp. 43-48, 1993.

[16] S. H. R Herman, 'Load Models for Mixed-class Domestic and Fixed, Constant Power Loads for use in Probabilistic LV Feeder analysis', *Electric Power Systems Research,* vol. 66, no. 2, pp. 149 – 153, 2003.

[17] E. Hirst, W. Lin, and J. Cope, 'Residential energy use model sensitive to demographic, economic, and technological factors', *Journal Name: Q. Rev. Econ. Bus.; (United States); Journal Volume: 17:2,* pp. Medium: X; Size: Pages: 7-22.

[18] G. P. Saha, and J. Stephenson, 'A model of residential energy use in New Zealand', *Energy,* vol. 5, no. 2, pp. 167-175, 1980.

[19] R. Nesbakken, 'Price sensitivity of residential energy consumption in Norway', *Energy Economics,* vol. 21, no. 6, pp. 493-515, 1999.

[20] J. Bentzen, and T. Engsted, 'A revival of the autoregressive distributed lag model in estimating energy demand relationships', *Energy,* vol. 26, no. 1, pp. 45-55, 2001.

[21] H. K. Ozturk, O. E. Canyurt, A. Hepbasli, and Z. Utlu, 'Residential-commercial energy input estimation based on genetic algorithm (GA) approaches: an application of Turkey', *Energy and Buildings*, vol. 36, no. 2, pp. 175-183, 2004.

[22] D. Young, 'When do energy-efficient appliances generate energy savings? Some evidence from Canada', *Energy Policy*, vol. 36, no. 1, pp. 34-46, 2008.

[23] X. Labandeira, Labeaga Azcona, J. Maria and Rodríguez Méndez, Miguel, , 'A Residential Energy Demand System for Spain ', *The Energy Journal* vol. 27, no. 2, pp. 87–111, 2006.

[24] T. Siller, M. Kost, and D. Imboden, 'Long-term energy savings and greenhouse gas emission reductions in the Swiss residential sector', *Energy Policy*, vol. 35, no. 1, pp. 529-539, 2007.

[25] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi, 'A bottom-up approach to residential load modeling', *Power Systems, IEEE Transactions on*, vol. 9, no. 2, pp. 957-964, 1994.

[26] R. Yao, and K. Steemers, 'A method of formulating energy load profile for domestic buildings in the UK', *Energy and Buildings*, vol. 37, no. 6, pp. 663-671, 2005.

[27] M. Stokes, 'Removing barriers to embedded generation : a fine-grained load model to support low voltage network performance analysis', Institute of Energy and Sustainable Development, DeMontfort University . Leicester, 2005.

[28] J. V. a. L. Paatero, P. D., 'A model for generating household electricity load profiles', *International Journal of Energy Research*, vol. 30, no. 5, pp. 273-290, 2006.

[29] M. S. M. Armstrong, H. Ribberink, I. Beausoleil-Morrison, J. Millette, 'Synthetically derived profiles for representing occupant-driven electric loads in Canadian housing', *Journal of Building Performance Simulation*, vol. 2, no. 1, pp. 15-30, 2009.

[30] I. Richardson, M. Thomson, D. Infield, and C. Clifford, 'Domestic electricity use: A high-resolution energy demand model', *Energy and Buildings*, vol. 42, no. 10, pp. 1878-1887, 2010.

[31] F. McLoughlin, A. Duffy, and M. Conlon, 'Evaluation of time series techniques to characterise domestic electricity demand', *Energy*, vol. 50, no. 0, pp. 120-130, 2013.

[32] D. J. Leith, M. Heidl, and J. V. Ringwood, 'Gaussian process prior models for electrical load forecasting', in Probabilistic Methods Applied to Power Systems, 2004 International Conference on, 2004, pp. 112-117.

[33] J. Lourenco, and P. Santos, 'Short term load forecasting using Gaussian process models', *Coimbra: Instituto de Engenharia de Sistemas e Comput adores de*, pp. 1-12, 2010.

[34] R. Singh, B. C. Pal, and R. A. Jabr, 'Statistical Representation of Distribution System Loads Using Gaussian Mixture Model', *Power Systems, IEEE Transactions on*, vol. 25, no. 1, pp. 29-37, 2010.

[35] A. O'Hagan, and J. F. C. Kingman, 'Curve Fitting and Optimal Design for Prediction', *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 40, no. 1, pp. 1-42, 1978.

[36] A. G. Riddell, and K. Manson, 'Parametrisation of domestic load profiles', *Applied Energy*, vol. 54, no. 3 SPEC. ISS., pp. 199-210, 1996.

[37] S. P. Moutter, P. S. Bodger, and P. T. Gough, 'Spectral decomposition and extrapolation of variations in electriciy loading', *IEE Proceedings C: Generation Transmission and Distribution*, vol. 133, no. 5 pt C, pp. 247-255, 1986.

[38] E. González-Romera, M. A. Jaramillo-Morán, and D. Carmona-Fernández, 'Monthly electric energy demand forecasting with neural networks and Fourier series', *Energy Conversion and Management*, vol. 49, no. 11, pp. 3135-3142, 2008.

[39] A. Graps, 'An introduction to wavelets', *Computational Science & Engineering, IEEE*, vol. 2, no. 2, pp. 50-61, 1995.

[40] H. Chen, C. A. Canizares, and A. Singh, *ANN-based short-term load forecasting in electricity markets*, vol. 2, pp. 411-415, 2001.

[41] J. V. Ringwood, D. Bofelli, and F. T. Murray, 'Forecasting electricity demand on short, medium and long time scales using neural networks', *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 31, no. 1-3, pp. 129-147, 2001.

[42] K. J. Krarti M, Cohen D, Curtiss P, 'Estimation of energy saving for building retrofits using neural networks', *Journal of Solar Energy Engineering*, vol. 120, pp. 211-216, 1998.

[43] F. I. Issa RRA, Asmus M, 'Development of a neural network to predict residential energy consumption', Proceedings of the sixth international

conference on the application of artificial intelligence to civil & structural engineering computing, 2001, pp. 65-66.

[44] M. Anstett, and J. F. Kreider, 'Application of neural networking models to predict energy use', *ASHRAE Transactions*, vol. 99, no. pt 1, pp. 505-517, 1993.

[45] J. F. Kreider, and X. A. Wang, 'Artificial neural networks demonstration for automated generation of energy use predictors for commercial buildings', *ASHRAE Transactions,* vol. 97, no. 1, pp. 775-779, 1991.

[46] J. F. Kreider, D. E. Claridge, P. Curtiss, R. Dodier, J. S. Haberl, and M. Krarti, 'Building energy use prediction and system identification using recurrent neural networks', *Journal of Solar Energy Engineering, Transactions of the ASME,* vol. 117, no. 3, pp. 161-166, 1995.

[47] M. Krarti, J. F. Kreider, D. Cohen, and P. Curtiss, 'Estimation of energy savings for building retrofits using neural networks', *Journal of Solar Energy Engineering, Transactions of the ASME,* vol. 120, no. 3, pp. 211-216, 1998.

[48] J. F. Kreider, and X. A. Wang, 'Improved artificial neural networks for commercial building energy use prediction', *Solar Engineering ASM*, vol. 1, pp. 361-366, 1992.

[49] D. A. Cohen, and M. Krarti, 'A neural network modeling approach applied to energy conservation retrofits', *Proceedings of the Building Simulation Fourth International Conference*, pp. 423-430, 1995.

[50] J. F. Kreider, and J. S. Haberl, 'Predicting hourly building energy use: the great energy predictor shootout - overview and discussion of results', *ASHRAE Transactions,* vol. 100, no. 2, pp. 1104-1118, 1994.

[51] M. Aydinalp-Koksal, and V. I. Ugursal, 'Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector', *Applied Energy,* vol. 85, no. 4, pp. 271-296, 2008.

[52] M. Aydinalp, V. Ismet Ugursal, and A. S. Fung, 'Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks', *Applied Energy,* vol. 71, no. 2, pp. 87-110, 2002.

[53] S. MohammadZadeh, and A. A. Masoumi, 'Modeling residential electricity demand using neural network and econometrics approaches', in Computers and Industrial Engineering (CIE), 2010 40th International Conference on, 2010, pp. 1-6.

[54] M. Aydinalp, V. I. Ugursal, and A. S. Fung, 'Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks', *Applied Energy,* vol. 79, no. 2, pp. 159-178, 2004.

[55] H. Mori, and H. Kobayashi, 'Optimal fuzzy inference for short-term load forecasting', *IEEE Transactions on Power Systems,* vol. 11, no. 1, pp. 390-396, 1996.

[56] Y. Y. Hsu, and K. L. Ho, 'Fuzzy expert systems: an application to short-term load forecasting', *Generation, Transmission and Distribution, IEE Proceedings C,* vol. 139, no. 6, pp. 471-477, 1992.

[57] P. A. Mastorocostas, J. B. Theocharis, and A. G. Bakirtzis, 'Fuzzy modeling for short term load forecasting using the orthogonal least squares method', *IEEE Transactions on Power Systems,* vol. 14, no. 1, pp. 29-36, 1999.

[58] H. Xu, and T. Niimura, 'Short-term electricity price modeling and forecasting using wavelets and multivariate time series', in Power Systems Conference and Exposition, 2004. IEEE PES, 2004, pp. 208-212 vol.1.

[59] Y. Chen, P. B. Luh, C. Guan, Y. Zhao, L. D. Michel, M. A. Coolbeth, P. B. Friedland, and S. J. Rourke, 'Short-term load forecasting: Similar day-based wavelet neural networks', *IEEE Transactions on Power Systems,* vol. 25, no. 1, pp. 322-330, 2010.

[60] J. Pahasa, and N. Theera-Umpon, 'Short-term load forecasting using wavelet transform and support vector machines', in Power Engineering Conference, 2007. IPEC 2007. International, 2007, pp. 47-52.

[61] H. T. Nguyen, and I. T. Nabney, 'Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models', *Energy,* vol. 35, no. 9, pp. 3674-3685, 2010.

[62] O. Ardakanian, S. Keshav, and C. Rosenberg, *Markovian models for home electricity consumption*, pp. 31-36, 2011.

[63] F. McLoughlin, A. Duffy, and M. Conlon, 'The generation of domestic electricity load profiles through Markov chain modelling', *Euro-Asian Journal of Sustainable Energy and Development Policy*, vol. 3, 2010.

[64] E. Carpaneto, and G. Chicco, 'Probabilistic characterisation of the aggregated residential load patterns', *IET Generation, Transmission and Distribution*, vol. 2, no. 3, pp. 373-382, 2008.

[65] J. L. Harris, and L.-M. Liu, 'Dynamic structural analysis and forecasting of residential electricity consumption', *International Journal of Forecasting*, vol. 9, no. 4, pp. 437-455, 1993.

[66] M. Ranjan, and V. K. Jain, 'Modelling of electrical energy consumption in Delhi', *Energy*, vol. 24, no. 4, pp. 351-361, 1999.

[67] H. A. Amarawickrama, and L. C. Hunt, 'Electricity demand for Sri Lanka: A time series analysis', *Energy*, vol. 33, no. 5, pp. 724-739, 2008.

[68] A. Cagni, E. Carpaneto, G. Chicco, and R. Napoli, 'Characterisation of the aggregated load patterns for extraurban residential customer groups', in Electrotechnical Conference, 2004. MELECON 2004. Proceedings of the 12th IEEE Mediterranean, 2004, pp. 951-954 Vol.3.

[69] V. Bianco, O. Manca, and S. Nardini, 'Electricity consumption forecasting in Italy using linear regression models', *Energy*, vol. 34, no. 9, pp. 1413-1421, 2009.

[70] A. J. R. Reis, and A. P. Alves da Silva, 'Feature extraction via multiresolution analysis for short-term load forecasting', *Power Systems, IEEE Transactions on*, vol. 20, no. 1, pp. 189-198, 2005.

[71] U. V. Swan L, Beausoleil-Morrison 'A new hybrid end-use energy and emissions model of the Canadian housing stock', in COBEE conference, Dalian, China, 2008.

[72] J. Boggis, 'Diversity, bias and balance', *Distribution of Electricity*, pp. 357–362, July, 1953.

[73] D. H. O. McQueen, P. R. Hyland, and S. J. Watson, 'Monte Carlo simulation of residential electricity demand for forecasting maximum demand on distribution networks', *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1685-1689, 2004.

[74] C. S. Chen, M. S. Kang, J. C. Hwang, and C. W. Huang, 'Temperature effect to distribution system load profiles and feeder losses', *Power Systems, IEEE Transactions on*, vol. 16, no. 4, pp. 916-921, 2001.

[75] F. Pilo, G. Celli, S. Mocci, and G. G. Soma, 'Multi-objective programming for optimal DG integration in active distribution systems', in Power and Energy Society General Meeting, 2010 IEEE, 2010, pp. 1-7.

[76] F. P. G. Celli, G.G. Soma, 'A Multi-Objective approach to investigate active distribution network impact on the constraining goals of the distribution system stakeholders', in Proceeding of CIGRE, General Session, Paris, 2010.

[77] E. G. G. Celli, S.Mocci, F. Pilo, G. Pisano, G.G. Soma, 'Reliability assessment of active distribution networks', in Proc. of CIGRE' General Session 2010, Paris.

[78] S. M. G. Celli, F. Pilo, M. Loddo, , 'Optimal Integration of Energy Storage in Distribution Networks', in Proc. of PowerTech conference, conference, Bucharest, Romania, 28 June – 2 July.2009

[79] F. P. G. Celli, G.G. Soma, M. Gallanti, R. Cicoria, 'Active Distribution Network Cost/Benefit Analysis with Multi-Objective Programming', in Proc. of CIRED conference, Praga, 2009.

[80] E. G. G. Celli, S. Mocci, F. Pilo, G. Pisano, G.G. Soma, 'From Passive to Active Distribution Networks: Methods and Models for Planning Network Transition and Development', in Proc. of CIGRE'General Session, Paris, 2008.

[81] G. Carpinelli, G. Celli, S. Mocci, F. Pilo, and A. Russo, 'Optimisation of embedded generation sizing and siting by using a double trade-off method', *Generation, Transmission and Distribution, IEE Proceedings-*, vol. 152, no. 4, pp. 503-513, 2005.

[82] G. Celli, E. Ghiani, S. Mocci, and F. Pilo, 'A multiobjective evolutionary algorithm for the sizing and siting of distributed generation', *Power Systems, IEEE Transactions on*, vol. 20, no. 2, pp. 750-757, 2005.

[83] G. C. G. Carpinelli, F. Pilo, A. Russo, 'Embedded Generation Planning under Uncertainty including Power Quality Issues', *European Transactions on Electrical Power*, vol. 13, no. 6, pp. 381–389, November/December, 2003.

[84] G. C. V. Allegranza, R. Cicoria, S. Mocci, F. Pilo, 'An Integrated Tool for Optimal Active Network Planning', in Proc. of CIRED conference, Barcelona, Spain, 2003.

[85] R. C. Celli, S. Mocci, F. Pilo, 'Probabilistic Optimization of MV Distribution Network in presence of Distributed Generation', in Proc. of 14th PSCC Conference, Sevilla, Spain, 2002.

[86] G. C. G. Carpinelli, F. Pilo, A. Russo, 'Distributed Generation Siting and Sizing under Uncertainty', in Proc. of Power Tech Conference, Porto, Portugal, 2001

[87] F. P. G. Celli, 'Optimal Distributed Generation Allocation in MV Distribution Networks', in Proc. of 22nd PICA Conference, Sydney, Australia, 2001, pp. 81-86.

[88] G. Celli, and F. Pilo, 'Optimal sectionalizing switches allocation in distribution networks', *Power Delivery, IEEE Transactions on*, vol. 14, no. 3, pp. 1167-1172, 1999.

[89] R. H. CT Gaunt, H Kadada, 'Design parameters for LV feeders to meet regulatory limits of voltage magnitude', in International Conference on Electricity Distribution (CIRED), Frankfurt, 2011.

[90] R. H. CT Gaunt, 'Impact of HWC DSM on voltage conditions and the choice of load parameters for feeder design', in 6th Cigré Southern Africa Regional Conference, Somerset West, August, 2009.

[91] R. G. Herman, C.T, 'Modeling and forecasting residential loads as probabilistic currents for LV network design', in IASTED PES, Clearwater, FL, USA, 2007 pp. 142-146.

[92] S. H. CG Carter-Brown, CT Gaunt, 'Distribution technical loss calculation and costing for network planning and design', in International Conference on Electricity Distribution (CIRED), Turin, 2005.

[93] C. G. IA Ferguson, 'LV network sizing in electrification projects - Replacing a deterministic method with a statistical method', in International Conference on Electricity Distribution (CIRED), Barcelona, 2003.

[94] R. H. CT Gaunt, M Dekenah, RL Sellick, SW Heunis, 'Data collection, load modelling and probabilistic analysis for LV domestic electrification', in International Conference on Electricity Distribution (CIRED), Nice, 1999.

[95] C. G. R Herman, SW Heunis, 'Benchmark tests and results for the evaluation of LV distribution voltage drop calculation procedures', *Transactions SA Institute of Electrical Engineers*, vol. 90, no. 2, pp. 54, June, 1999.

[96] C. G. R Herman, SW Heunis, 'The effect of voltage drop of connecting customers to L.V. feeders in different configurations', *Transactions of the South African Institute of Electrical Engineers*, vol. 89, no. 1, pp. 27-32, March, 1998.

[97] S. H. R Herman, 'A general probabilistic voltage drop calculation method for LV distribution networks based on a Beta pdf load model', *Electric Power Systems Research*, vol. 46, pp. 45-49, 1998.

[98] J. M. R Herman, 'Voltage regulation algorithm for a bi-phase distribution system feedingresidential customers using a Beta pdf load model', *Electric Power Systems Research*, vol. 43, no. 2, pp. 77-80, 1997.

[99] C. G. RL Sellick, 'Comparing methods of calculating voltage drop in low voltage feeders', *Transactions SA Institute of Electrical Engineers*, vol. 86, no. 3, September, 1995.

[100] J. M. R Herman, JHR Enslin, 'The analysis of voltage regulation using the beta distribution model', *Electric Power Systems Research*, vol. 29, pp. 213 - 216., 1994.

[101] F. Demailly, O. Ninet, and A. Even, 'Numerical tools and models for Monte Carlo studies of the influence on embedded generation on voltage limits in LV grids', *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp. 2343-2350, 2005.

[102] M. Davies, and R. G. Paterson, 'Phase unbalance in low-voltage distribution systems', *Proceedings of the IEE - Part A: Power Engineering*, vol. 109, no. 48, pp. 535-541, 1962.

[103] 'Federal Solar Cities Initiative'; http://www.climatechange.gov.au/solarcities, accessed March 10th 2013.

[104] 'Alice Solar City'; http://www.alicesolarcity.com.au/, accessed March 10th 2013.

[105] 'Perth Solar City'; http://www.perthsolarcity.com.au/, accessed March 10th 2013.

Chapter 3 A refined distribution feeder classification method

In this chapter the mathematical foundations of cluster analysis for determining the prototypical distribution feeder models are established. The method made use of statistical cluster analysis and discriminant analysis.

3.1 Distribution feeder taxonomy using hierarchical cluster analysis

3.1.1 Cluster analysis principles

Clustering is the process of dividing data elements into classes or clusters, so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible [1]. It is an important exploratory tool widely used in many areas such as biology, sociology, medicine and business. For example, cluster analysis was used to detect classes or sub-classes of diseases in biology.

Cluster analysis is commonly used to categorise objects into homogeneous groups based upon a vector of variables. In our study, feeder data will be clustered into smaller disjoint sets $c_i(i=1,...,k)$, and k is the number of clusters. The feeder data Z are composed of vectors z_i of n variables, u_j [2]:

$$Z = \{z_i | z_i = (u_j), i = 1,...,k; j = 1,...,n\} \tag{3.1}$$

$$c_i \neq 0, i = 1,...,k \tag{3.2}$$

$$\bigcup_{i=1}^{m} c_i = Z \tag{3.3}$$

$$c_i \bigcap c_j = \varnothing, \ \ i \neq j, i = 1,...,k, j = 1,...,k \tag{3.4}$$

Clustering follows the rules that each cluster contains at least one feeder, no feeder lost during the cluster process, and each feeder can only belong to one certain cluster.

A prototypical distribution feeder has a number of parameters, such as feeder length and feeder load, that lie at the centroid of a cluster of feeders. For a given existing physical feeder it is possible to measure a geometric distance or Euclidean norm that expresses the closeness of this feeder to the prototypical feeder. The principle of feeder clustering and prototypical feeder selection is shown in Figure 3.1.

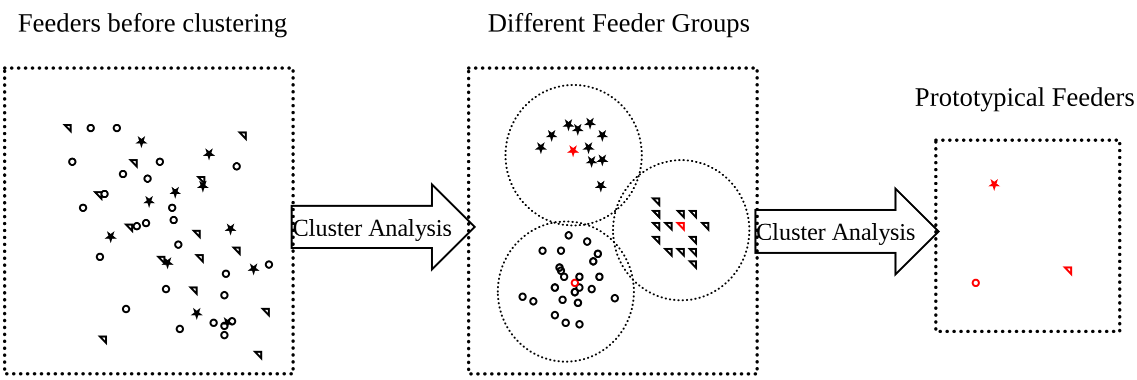


Figure 3.1 Graph explains the principle of statistical feeder cluster analysis

3.1.2 Procedures of cluster analysis

Generally, cluster analysis involves seven main procedures as summarised by Milligan [3]:

1) Selection of clustering elements: the sample should represent the population. One of the strategies is to remove outliers from the dataset first, and the relationship between the established clusters and the outliers can be back tested afterwards.
2) Selection of the differentiating variables: it is necessary to select enough variables to provide sufficient information, however, only those variables that believed to be

helpful in discriminating between members of different the clusters should be chosen.

3) Data standardisation: Milligan and Cooper [4] simulated and compared the clustering using eight forms of standardisation, including the unstandardized data. Care needs to be taken whether the clustering should be based on the raw or the standardised variable space.
4) Choosing a measure of association (dissimilarity/similarity measure): the measure should reflect those characteristics that are suspected to distinguish the clusters present in the data. Commonly used techniques cluster elements based on the dissimilarity matrix. The degree of dissimilarity between elements is usually measured by their distance.
5) Selection of clustering method: clustering methods should be efficient in recovering the cluster structure underlying the data.
6) Determining the number of clusters: it is necessary to estimate the number of clusters since most clustering methods partition objects into a fixed number of clusters.
7) Interpretation: the presence of a single ideal type in each cluster would be a way to interpretation for the cluster. Graphical representations could be very helpful.

This thesis applies Milligan's basic principles for cluster analysis with specific adaptations for the study of distribution feeders. The thesis additionally applies discriminant analysis as a further refining set.

I) Element selection

The MV and LV distribution feeder data were collected by Western Power Corporation. A data cleaning process first removed feeders where the data sourced from several databases had any inconsistencies.

II) Variable selection

The selection of the clustering variables is crucial to statistical analysis. Tens of parameters can be used to characterise distribution feeders. Many algorithms perform poorly in a high dimensional space. Some variables just contribute 'noise' and interfere with clustering process. This difficulty in analysing datasets with many variables and a small number of samples is known as the curse of dimensionality [5]. Reducing the number of variables before clustering can avoid such difficulty. Also, it reduces the computing time, and models established in that way are easier to be interpreted. The main disadvantage of dimensionality reduction is the possible risk of information loss.

The selection of the clustering input variables, u_j, is made on an iterative basis in this thesis. Variables are selected that are expected to differentiate the feeders and clusters are formed. The variable selection is then back tested using analysis of variance (ANOVA) tests. The level of significance of a variable is measured by the F ratio:

$$F = \frac{SSE/(k-1)}{T/(j-k)} \tag{3.5}$$

F follows the distribution $F[(k-1),(j-k)]$ of Fisher-Snedecor with $(k-1)$ and $(j-k)$ degree of freedom, and the significance level of a variable is defined as [6]:

$$Sig. = Prob(F(k-1),(j-k) > F) \tag{3.6}$$

The significance level of a certain variable gives information regarding whether the variable is discriminative and meaningful in the analysis. Variables with larger F values or smaller *Sig.* values contribute most strongly to the cluster solution. Details and examples on the parameter selections for the MV and LV feeder sets will be given in the next chaper.

III) Data standardisation

Feeder data were normalised by *Z*-score transformation. The *Z*-score transformation standardises variables to the same scale, producing a set of variables with a mean of 0 and a standard deviation of 1.

IV) Measure of association

The clustering was based on the dissimilarity matrix with the degree of dissimilarity measured by squared Euclidean distance (*SEUCID*). *SEUCID* is the sum of the squared differences between the values for the items. It is used to determine the separation of feeders within the clustering space. The separation of feeders *x* and *y* is expressed as:

$$SEUCID(x,y) = \sum_{i=1}^{q} (x_i - y_i)^2 \qquad (3.7)$$

where, *q* is dimensionality of clustering space, i.e. the number of variables (*q* = 7).

V) Clustering method

Clustering methods can be nonhierarchical or hierarchical. A major advantage of hierarchical methods over nonhierarchical methods is that the number of clusters does not need to be known or set. Nonhierarchical methods may only be used to cluster items and could not determine the optimal number of clusters.

Hierarchical methods include agglomerative hierarchical methods and divisive hierarchical methods. Agglomerative hierarchical methods start with clusters containing one object and combines objects until all objects form only one cluster [7, 8]; while divisive hierarchical methods start with one cluster and split objects successively to form clusters.

Agglomerative hierarchical methods have been used in this thesis. The main steps for the agglomerative hierarchical clustering algorithm are as the follows.

- To start with, each element formed one cluster.
- The dissimilarities were calculated in order to search for the most similar feeder pair *r* and *s*.
- Feeders *r* and *s* were combined into a new cluster that reduced the number of clusters by 1. The dissimilarities were calculated again between the newly formed cluster and all remaining clusters.
- The second and third steps were repeated for $(n-1)$ times until all objects combined into one cluster.

Hierarchical clustering by Ward’s method was chosen to categorise the MV and LV distribution feeders. Ward's method, or more precisely Ward's minimum variance method, was originally presented by J. H. Ward [9]. It minimises the total within-cluster variance. Each feeder was firstly allocated to be a cluster on its own, so the number of clusters to begin with is equal to the number of feeders. A dissimilarity measure of Euclidean distance was used, which assumes the larger the distance the less similar of the two elements. At each step, Euclidean distance between clusters is calculated to find out similarities between clusters, and the pair of clusters with minimum cluster distance are merged [8]. The algorithm proceeds iteratively, at each stage joining the two most similar clusters, eventually until there is only one cluster remains.

VI) Optimal number of clusters

A crucial step was to determine the optimal number of clusters (m_{opt}) by analysing the three statistical parameters [10-12]:

- Semi-Partial R^2;
- Pseudo- F statistic and
- Pseudo-T^2 Test.

These parameters are observed during a progressive clustering activity which commences with the number of clusters equal to the number of data points and terminates with a single cluster.

A) Semi-Partial R^2

Semi-Partial R^2 *(SPRSQ)* measures the loss of between-cluster distance caused by grouping two clusters, or the decrease in R^2. R^2 for k clusters is defined as [7]:

$$R_k^2 = \frac{T - \sum_k SSE_k}{T} \tag{3.8}$$

where, k is the number of clusters. Given n objects with p variables, the sum of squares within clusters where each object forms its own group is zero. For all objects in a single group, the sum of squares for error (*SSE*) is equal to the total sum of squares:

$$SSE = \sum_{i=1}^{m} \left(y_i - \overline{y}\right)' \left(y_i - \overline{y}\right) = \sum_{i=1}^{m} \left\| y_i - \overline{y} \right\|^2 = T \tag{3.9}$$

Thus, the sum of squares within clusters is between zero and *SSE*. SSE_k is the internal cluster distance for cluster *k*. *T* is the total distance for *m* clusters:

$$T = \sum_{i=1}^{m} \left\| y_i - \overline{y} \right\|^2 \tag{3.10}$$

Joining two clusters, the *SPRSQ* describes the decrease of R^2 when moving from *k* to (*k-1*) clusters, and is defined as:

$$SPRSQ = R_k^2 - R_{k-1}^2 \tag{3.11}$$

The larger the value of *SPRSQ*, the more separated the clusters. Small magnitudes for the Semi-Partial R^2 parameter indicates that the two clusters can be regarded as identical. Ideally, cluster number selection is made to maintain the smallest value; however, a balance is needed between error level and number of clusters.

B) Pseudo- *F* statistic

The pseudo-*F* (*PSF*) Statistic measures the separation among the clusters at the current level in the hierarchy, and is defined as [7]:

$$F_k^* = \frac{\left(T - \sum_k SSE_k \,/(k-1)\right)}{\sum_k SSE_k \,/(n-k)} \tag{3.12}$$

High *PSF* values indicate that the mean vectors of all clusters are different. High values followed by lower values indicate possible cluster solutions. Therefore, cluster selection can be made around peaks in *PSF* [13].

C) Pseudo-T^2 Test

The pseudo-T^2 (PST^2) Test is the ratio of the increased within-category variance when joining two clusters *r* and *s* each having n_r and n_s elements, to the variance within each of the two clusters, and is defined as [7]:

$$PST^2 = \frac{[SSE_t - (SSE_r + SSE_s)](n_r - n_s - 2)}{SSE_r + SSE_s} \tag{3.13}$$

Small PST^2 indicates that clusters can safely be combined. On the other hand, high PST^2 values mean high differences in the mean vectors of the two clusters just combined [14].

VII) Interpretation of cluster analysis results

The traditional representation of the cluster hierarchy is a dendrogram or tree graph, where all elements to be clustered at one end and clusters at the other. The *x*-axis illustrates the distance at which the clusters merge, while the objects are placed along the *y*-axis. Cutting the tree graph at a given height offers a clustering result at a particular accuracy level.

For the purposes of analysis, for example the assessment of a technology upon a physical feeder, the cluster information is used to identify an existing physical feeder. The representative feeder is that physical feeder with parameters with a minimum cluster center. In each cluster, the feeder with the smallest value of *SEUCID* was picked to be the representative one in that particular cluster. The clustering approach produced clusters with known centroids in terms of the clustering variables. An example of feeder grouping using cluster analysis will be given in Section 3.3.

3.2 Distribution feeder taxonomy using discriminant analysis

One of the disadvantages of hierarchical single-pass clustering is that it cannot correct later for erroneous decisions made earlier. In this section an improvement upon it is presented.

3.2.1 Principle of discriminant analysis (DA)

Discriminant analysis has widespread applications in scientific and engineering areas, such as economy, social sciences and health sciences. Discriminant function (DF) analysis is used to determine which variables discriminate between two or more naturally occurring groups [15]. A classifier can be constructed using discriminant functions of the input variables to determine to which group an individual case belongs.

Discriminant analysis can provide an insight into the information entropy of the selected variables. A contribution of the work was to demonstrate successful clustering, or classification, using a small number of meaningful variables.

3.2.2 Process of the refined classification method

As shown in Figure 3.2, the key steps in the refined classification method start with feeder data acquisition followed by the data pre-treatment. The feeder data were collected within the SWIS by Western Power Corporation. The data were drawn from several databases and the pre-treatment was to remove any data with inconsistencies.

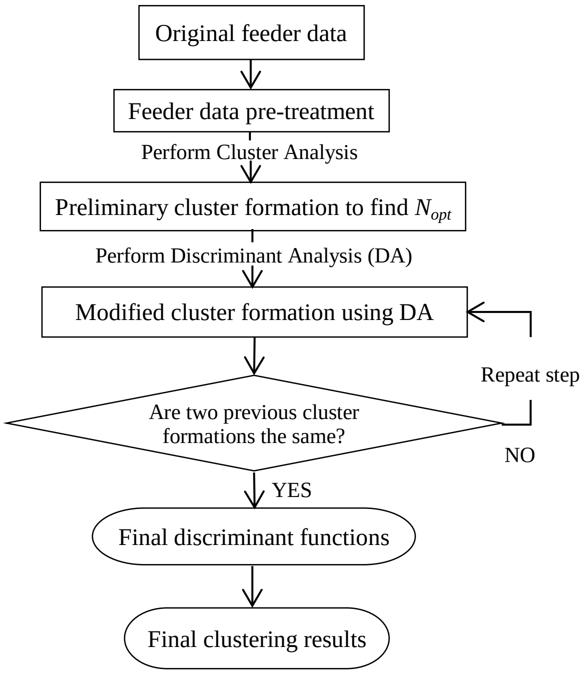


Figure 3.2 Flow chart used in the refined classification method

Cluster analysis gave a preliminary classification and an optimal cluster number m_{opt} which has been used as a starting point so that DA can classify the feeders into m_{opt} clusters based on Bayes rules [16]. After the initial DA process, it has been found out that the classification results were not the same as the ones from the primary cluster analysis, indicating the first cluster had some unstable features. Clustering is a hierarchical single-pass process. Optimal classification requires, in general, back-tracking and multiple-passes. The DF from the first round DA classification has to be refined. A second round of discriminant analysis was conducted based on the set of discriminant functions extracted from the first round of discriminant analysis. A comparison of the classification results to those found in the first DA round showed some changes. After seven iterative rounds of DF re-substitutions, the classification results from the seventh DA are exactly the same as from the sixth. The classification error rate is zero, showing stable cataloguing has been achieved.

DA and the discriminant function establishment requires pre-tests first to find out the validaty.

3.2.3 Pre-tests before building DF

The first step is to examine whether our case study is valid to undertake DA. If the DF are found to be statistically significant, then the groups can be distinguished based on predictor variables during DA. To start with, the method using Wilks' lambda for the significance testing, which can be extracted using statistical software packages such as SAS and SPSS [12, 15].

3.2.3.1 Wilks' lambda and multivariate *F* test

Wilks' lambda is multivariate generalization of the single variable *F*-distribution. It is a direct measure of the proportion of variance in the combination of dependant variables that is unaccounted for in the independent variable (grouping factor). Wilks' lambda varies from 0 to 1. The smaller the value of Wilks' lambda for a variable, the more it contributes to DF. Wilks' lambda follows the equation [17, 18]:

$$\Lambda = \left| \frac{SS_{wg}}{SS_{bg} + SS_{wg}} \right| \tag{3.14}$$

where, SS_{bg} and SS_{wg} are cross-products matrices for between-group differences and within-groups differences, respectively. Let λ_i be the ordered eigenvalues of $SS_{wg}^{-1} SS_{bg}$:

$$\Lambda = \prod_{i=1}^{n} \frac{1}{1+\lambda_i} \tag{3.15}$$

The level of significance of a variable is measured by the *F* ratio:

$$F = \frac{WSS/(k-1)}{TSS/(j-k)} \tag{3.16}$$

where, *WSS* is the within group sum of squares, and *TSS* is the total sum of squares.

The significance level of a certain variable gives information regarding whether the variable is discriminative and meaningful in the analysis. If the significance level value of each variable is very small, indicating the grouping variables are significantly different and will contribute strongly to the construction of the DF.

3.2.3.2 Tests for homogeneity of covariance matrices

Homogeneity of variance occurs where variance across multiple samples is similar. Homogeneity of variance-covariance matrices concerns the variance-covariance matrices of the multiple dependent measures for each group. Taking a data set with five dependent parameters for an example, it tests for five correlations and ten covariances for equality across the groups.

Tests for homogeneity of covariance matrices, i.e. homocedasticity, are essential in order to assess whether linear or quadratic discriminant functions are applicable in our case. The Box's *M* is used here for checking homocedasticity, which should be non-significant. The Box's *M* tests the statistical hypothesis that the variance-covariance matrices are equal. The Box's *M* can be written as [19, 20]:

$$M = (n-g)ln|S| - \sum_{i=1}^{g}(n_i - 1)ln|S_i| \tag{3.17}$$

where, n is the total sample size, g is the number of cells with non-singular covariance matrices, n_i is the number of cases in the i_{th} cell. In addition, S is pooled covariance matrix:

$$S = \sum_{i=1}^{g} n_i S_i / n \tag{3.18}$$

where, S_i is cell covariance matrix. When the level of significance is low, meaning the assumption of equality of covariances across groups does not apply in our case, the discriminant functions to be built should be quadratic rather than linear. The quadratic discriminant function is usually applied for the group classification where the covariance matrices in the populations are substantially unequal.

3.2.3.3 Tests of agreement and validation

When the significance level of each test of function is quite small, which proves that the set of DF are able to separate the feeder data in the statistical sense, and hence the functions are valid in the DA.

3.2.4 Discriminant functions

3.2.4.1 Quadratic discriminant functions

The beauty of discriminant analysis is that it can construct a set of discriminants for characterising group separation based upon grouping variables. Discriminants that are linear functions of the grouping variables are called linear discriminant functions (LDF). Linear discriminant analysis (LDA) constructs one or more LDF(s), i.e. linear combinations of the variables, such that the different groups differ as much as possible on the equations. The LDA is considered valid when the variance-covariance matrix of the variables is the same in all groups. However, if the assumption is violated, quadratic discriminant analysis is required.

3.2.4.2 Building Discriminant Functions

As it has been shown in the previous section, group means can be found to be statistically significant, then classification of variables is undertaken and quadratic DF can be built. The set of DF in one case study are independent. DA determines the optimal combination of variables, where the first function provides the most overall discrimination between groups, and the second provides second most, and so on. In other words, the first function picks up the most variation; the second function picks up the most significant part of the unexplained variation, and so on.

The quadratic discriminant function (QDF) can be expressed in the form [16]:

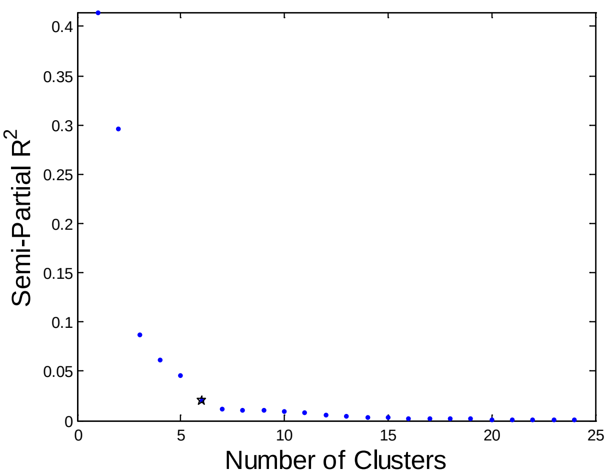
$$G_k(x) = a_0 + \sum_{i=1}^{d} a_i x_i + \sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij} x_i x_j \tag{3.19}$$

The feeders are classified into the groups, in which they had the highest classification scores. That is, by identifying the largest absolute correlations associated with each DF, the feeder under study can then be categorised into a particular cluster. An example of feeder classification using discriminant analysis will be given in Section 3.4.

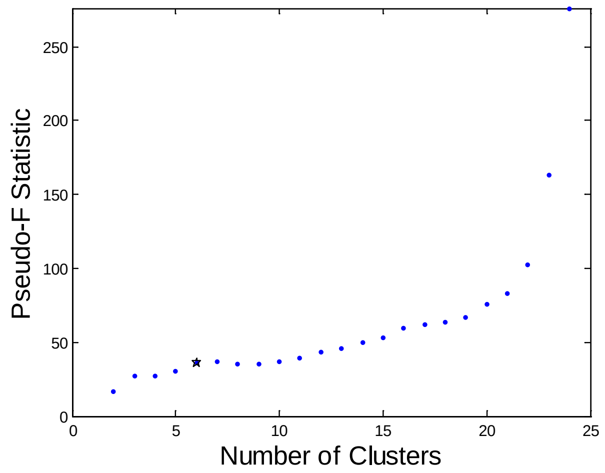
3.3 An example of clustering on a small LV feeder set

This feeder set with a smaller number of feeders was chosen as an example for cluster analysis for its clearer graphical demonstration. This example used data drawn from existing LV databases that are primarily established for distribution transformer load forecasting.

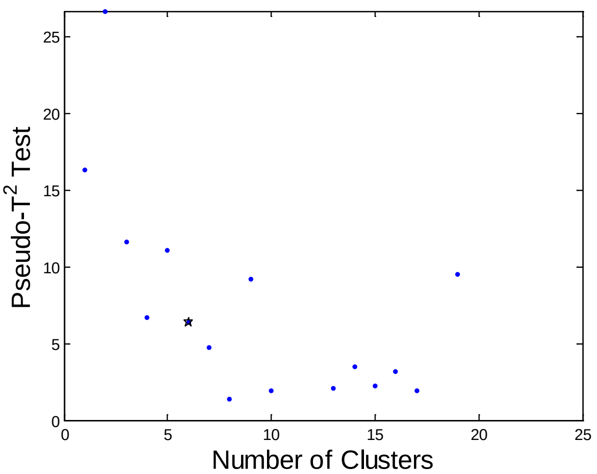
The LV networks is connected to one prototypical MV feeder type - metropolitan residential, named the NB 514.0 Reid Highway feeder, within the SWIS in Western Australia. The 25 LV feeders are to be clustered according to their seven key parameters, which will be explained in detail in the next chapter.



(a)



(b)



(c)

Figure 3.3 Graphs illustrate the method to determine the optimal cluster number, which is 9 (marked with ☆).

(a) – (c) showing the relationship between number of clusters and Semi-Partial R^2, Pseudo-F Statistic and Pseudo-T^2 Test, respectively.

After data standarisasion using Z-score transformation, optimal cluster numbers were identified for the 25 LV feeders. Figure 3.3 gives a set of graphs showing the method in determining optimal cluster numbers, which was 6 (m_{opt} = 6). Figure 3.3 (a) shows that the level of *SPRSQ* begins to increase rapidly when the allowable number of clusters is reduced to below six, so the favourable number of clusters could be no less than six according to the *SPRSQ* parameter. The pseudo-*F* parameter suggests six, seven or larger than ten. The pseudo-T^2 Test recommends three, four, six, eight, or ten. Therefore, from the above analysis, six clusters or subsets are selected to represent the set of LV feeders connected to the representative NB 514.0 Reid Highway MV feeder.

Hierarchical clustering, as shown in Figure 3.4, has been analysed for the 25 LV feeder set based upon the optimal number of clusters. Figure 3.4 demonstrates the dendrogram resulting from cluster analysis using the Ward's method. Outcomes of a dendrogram and the selection of the height or called cutting the tree have been illustrated in Figure 3.4, based on the LV feeder data. The optimal number of clusters decides the point where the tree graph to be cut or the suitable accuracy level. The optimal number of clusters was six, as mentioned in the previous section, so cutting the 'tree' at the point where Semi-partial *R*-squared is equal to 0.0206 will result in separating the 25 feeders into six subsets - each represents a different type of load or customer.

For the grouping process, it first starts with n = 25 clusters, just as there are 25 branches on the 'tree'. The 25 feeders are numbers from feeders DB 1 to DB 25. Combining DB 1 with DB 10 reduces in the number of clusters or branches by one, from 25 to 24. Joining feeder DB 8 with the newly combined cluster, reduces the number of clusters by one and results in 23 clusters. Then joining feeder DB 18 with the recently combined cluster, reduces the number of clusters by one again and results in 22 clusters. Finally, when all feeders are combined, they form a single cluster just like the root of the tree. Looking at the 'tree' from branches to root, i.e. from left to right, it can be seen that different groups of feeders are successively combined to form clusters. Once a feeder or group of feeders are combined they are never separated in the agglomerative clustering process.

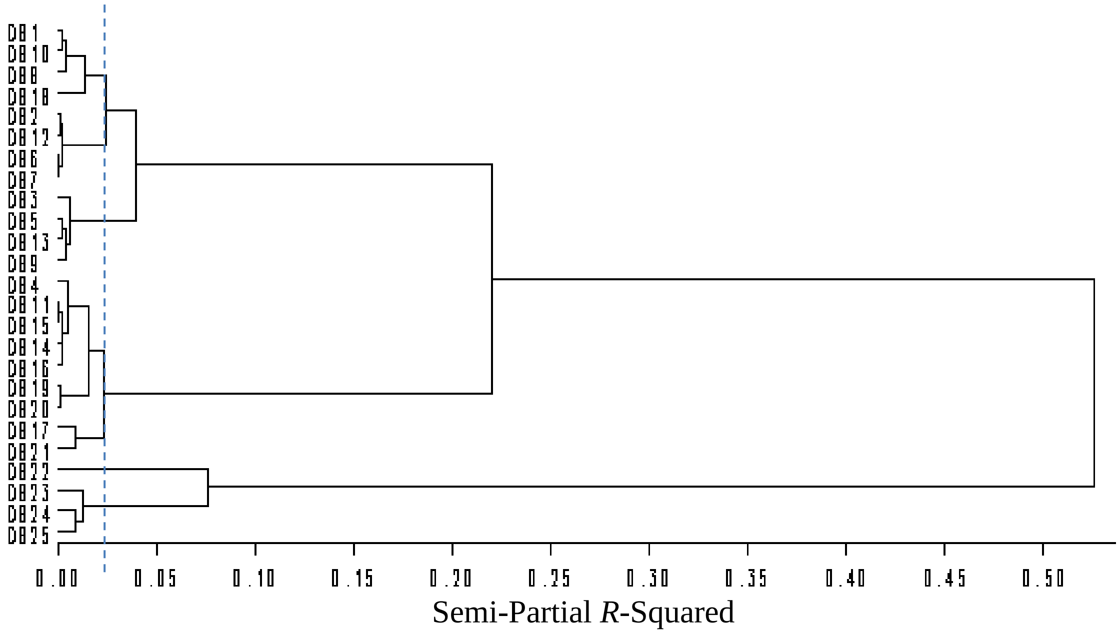


Figure 3.4 Dendrogram resulting from cluster analysis using the Ward's Method for the LV distribution feeders.

The results of the 25 LV feeder clustering are shown in Figure 3.5. The six clusters are designated with varying graphical symbols. The upper plot uses the three variables of underground feeder length (L_{UG}), total load capacity/rating ratio (R) and load capacity excluding residential customer (P_{others}). The lower plot uses variables of number of residential customer (N_r), load capacity of residential customer (P_r) and overhead feeder length (L_{OH}). These two figures show a high separation of the clusters in the three dimensional view.

It can be seen that Cluster 1 has very high values for load capacity for residential customers (P_r), numbers of residential customer and total load capacity/rating ratio (R), with zero overhead feeder length. This represents a type of residential user with high transformer utilisation using only underground lines. Cluster 5 has low values for numbers of residential customer and total load capacity/Rating ratio representing the feeders servicing commercial/Industrial consumers. Cluster 6 services business centre customers and has a high level of load capacity excluding residential customers, and low values of underground and overhead feeder length. Table 3.1 shows the characteristics of the six representative LV feeders among the 25 feeders.

Once the clustering process is complete it is possible to specify the mean parameters of each representative feeder type. Each physical feeder in the cluster can be ranked with respect to its variation from the mean parameters. In terms of Euclidean parameter distance there will be one physical feeder that is closest to the cluster centre.

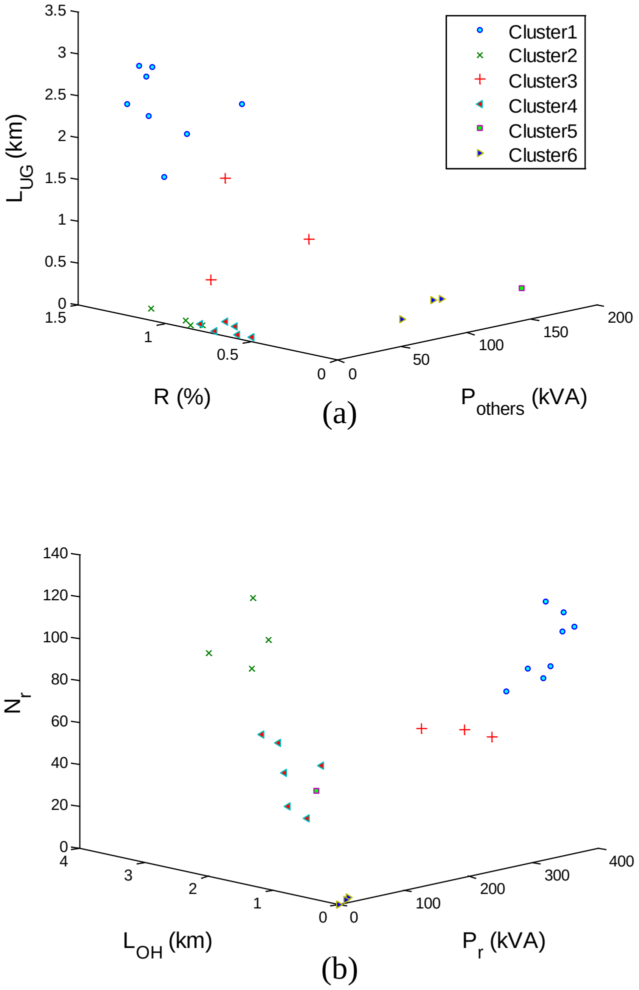


Figure 3.5 LV distribution feeder clustering on 7 key grouping variables (only 6 variables showing in Figure 3.5), with 6 symbols for the 6 clusters or subsets.

TABLE 3.1
Characteristics of the Six Representative LV Feeders Among the 25 Feeders

Cluster	**LV Feeder Name**	***SEUCID***	L_{OH} **(km)**	L_{UG} **(km)**	N_{others}	N_r	P_{others} **(kVA)**	P_r **(kVA)**	***R*** **(%)**
1	ALMADINE	0.1895	0.00	2.43	1	92	0.401	326.440	108.95
2	MELIA	0.1513	3.74	0.09	1	83	0.773	174.801	87.79
3	ANGLICAN HOMES 1	0.6379	0.00	0.61	2	64	4.458	236.578	76.52
4	LENNOXTOWN	0.1083	1.57	0.17	1	47	1.489	130.104	65.80
5	50 MARRI RD	0.0000	1.37	0.14	23	38	180.433	103.030	28.35
6	DUNCRAIG RECREATION CENTRE	0.9219	0.00	0.01	3	1	143.717	12.481	52.07

TABLE 3.2
Tests of Equality of Group Means

	Wilks' Lambda	***F***	**df1**	**df2**	**Sig.**
P_L **(Peak Load)**	0.303	78.0	8	271	0.000
N_C **(Customers)**	0.271	91.2	8	271	0.000
L_{MV} **(MV Length)**	0.202	133.8	8	271	0.000
L_{LV} **(LV Length)**	0.320	72.1	8	271	0.000
N_{TX} **(Number per transformer)**	0.200	135.1	8	271	0.000
C_{LV} **(Installed Capacity)**	0.163	173.8	8	271	0.000

(df: degree of freedom; Sig.: significance level; *F*: a test to identify which variable will most separate the group)

The next application step is to use the representative physical feeders as a basis for developing representative network models. The physical feeder data can be used to construct models at a level of detail suitable for the eventual purpose. These become test cases on which further research such as studies into DG or EV impacts or the application of power electronic devices or storage can be assessed. The outcomes of these studies can then be extrapolated to the larger network on the basis of the representative nature of the prototypical feeders.

3.4 An example of DA on a set of MV feeders

As an illustration, the DA will be carried out on a set of 280 MV distribution feeders within SWIS in this section.

A) Wilks' lambda and multivariate *F* test

As it can be seen from Table 3.2, the significance level of each variable is under 0.005, indicating the six feeder variables are significantly different and will contribute strongly to the construction of the DF.

B) Tests for homogeneity of covariance matrices

As it can be seen from Table 3.3, the level of significance is below 0.005, meaning the assumption of equality of covariances across groups does not apply in our case, so the discriminant functions to be built should be quadratic rather than linear. The quadratic discriminant function is usually applied for the group classification where the covariance matrices in the populations are substantially unequal.

TABLE 3.3
Box's *M* Test Results

Box's *M*		2507
F	Approx.	13.34
	df1	168
	df2	23309
	Sig.	.000
Tests null hypothesis of equal population covariance matrices.		

C) Tests of agreement and validation

As it can be seen from Table 3.4, the significance level of each test of function is less than 0.01, which proves the functions are valid in the DA.

TABLE 3.4
Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 6	0.002	1721	48	.000
2 through 6	0.018	1089	35	.000
3 through 6	0.108	604	24	.000
4 through 6	0.294	333	15	.000
5 through 6	0.722	88.6	8	.000
6	0.956	12.35	3	.006

D) Building quadratic discriminant functions

The resubstitution summary using QDF is given in Table 3.5. The rates are zero, indicating stable classification has obtained after seven iterations. The prior probabilities of group membership are equal. For a feeder classification with six variables, the first QDF can be written as:

$$
\begin{aligned}
G(x)_{k=1} &= a_0 + a_1x_1 + a_{11}x_1x_1 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4 + a_{15}x_1x_5 + a_{16}x_1x_6 \\
&+ a_2x_2 + a_{22}x_2x_2 + a_{23}x_2x_3 + a_{24}x_2x_4 + a_{25}x_2x_5 + a_{26}x_2x_6 \\
&+ a_3x_3 + a_{33}x_3x_3 + a_{34}x_3x_4 + a_{35}x_3x_5 + a_{36}x_3x_6 \\
&+ a_4x_4 + a_{44}x_4x_4 + a_{45}x_4x_5 + a_{46}x_4x_6 \\
&+ a_5x_5 + a_{55}x_5x_{\backslash 5} + a_{56}x_5x_6 \\
&+ a_6x_6 + a_{66}x_6x_6
\end{aligned}
$$

(3.20)

where, the coefficients of a_0, a_i and a_{ij} are listed in Table A.1 in Appendix. x_i (i = 1 to 6) are peak load P_L, number of customers N_C, MV circuit length $L_{MV,}$ LV circuit length L_{LV}, customers per transformer $N_{TX,}$ and installed capacity per customer C_{LV}, respectively. The details on the six variables will be given in the next chapter.

The first two items in the function are linear parts with *(d+1)* parameters, while the last item is quadratic part with *d(d+1)/2* additional parameters. The quadratic part is equal to zero if covariance matrices is homogeneous, meaning equality of covariances across groups.

E) Interpreting the discriminant functions

Any MV feeder in the 280 feeders has nine calculated value of *G (x)*, so it belongs to the group with the highest *G(x)* value. The summary of MV feeders classification is listed in Table 3.6, and the characteristics of the representative MV feeders in Perth are given in Table 3.7. Figure 3.6 shows the clustering outcome. Two three-dimensional plots are used to represent the six-dimensional space. In each cluster, the feeder with the smallest value of *SEUCID* was picked to be the representative one in that particular cluster.

TABLE 3.5
Resubstitution Summary Using Quadratic Discriminant Function

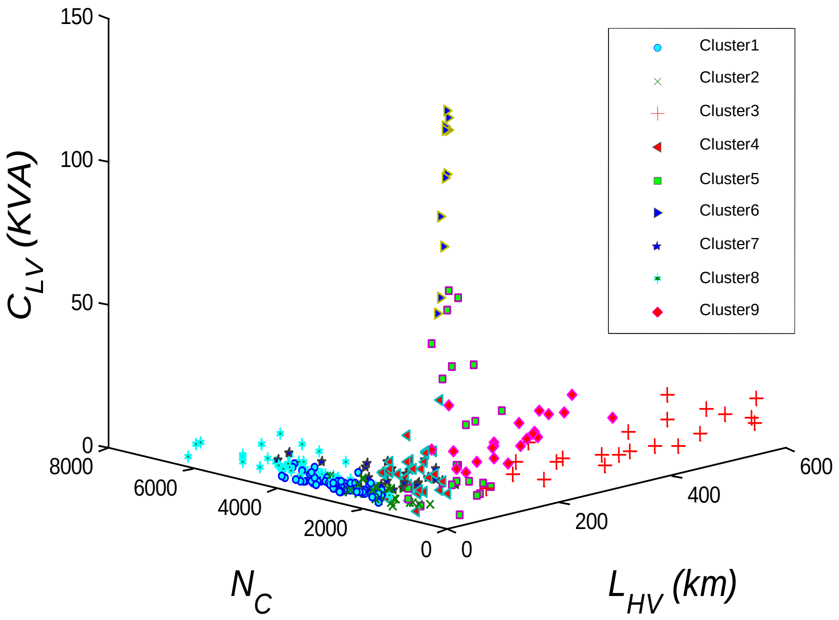
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Priors	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111

TABLE 3.6
Summary of MV Feeders Classification Using the Refined Method

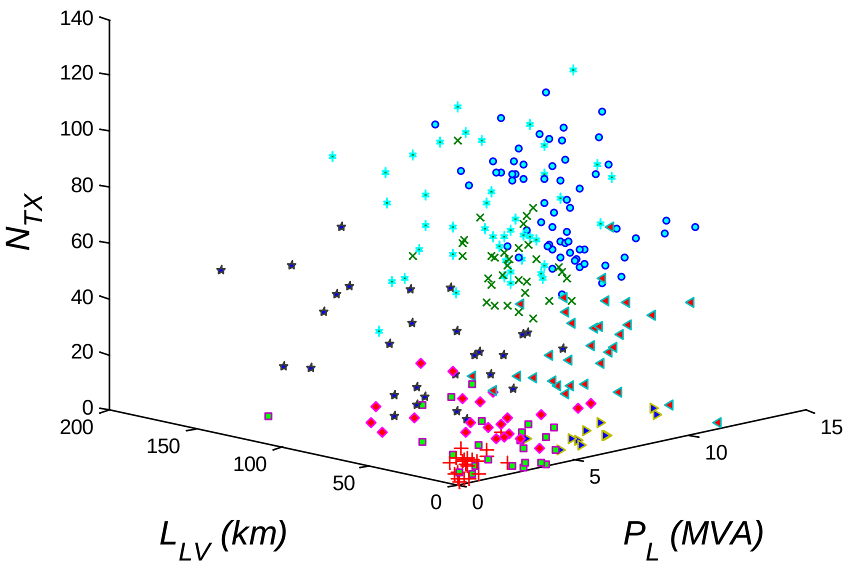
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
P_L **(Peak Load) MVA**	8.64	5.58	1.73	7.33	2.79	6.68	5.86	10.41	6.25
N_C **(Customers)**	2795	1915	659	913	410	156	2122	4085	1754
L_{MV} **(MV Length) km**	17.17	16.83	373.78	10.44	48.94	8.72	53.13	39.2	244.5
L_{LV} **(LV Length) km**	56.82	44.24	15.99	22.32	18.69	10.67	100.97	120.71	69.9
N_{TX} **(Number per transformer)**	77.17	63.38	2.11	33.67	7.58	5.26	26.1	61.82	4.22
C_{LV} **(Installed Capacity) kVA**	4.97	5.58	10.55	16	29.4	119.17	8.63	5.8	14.33
No. of Feeders	63	33	22	32	24	11	29	44	22

TABLE 3.7
Characteristics of the Nine Representative MV Feeders in Perth

Cluster	Feeder	Street	Substation	SEUCID	P_L (MVA)	N_C	L_{MV} (km)	L_{LV} (km)	N_{TX}	C_{LV} (kVA)
1	MLG 527.0 12 WIDGE RD	Widgee Rd	MALAGA	0.0286	9.35	2,995	18.8	59.4	74.9	4.4
2	NB 514.0 REID HWY	Reid Highway	NORTH BEACH	0.0792	5.89	1786	13.04	43.44	68.7	4.3
3	BTN 505.0 GWALIA STH	Gwalia South	BRIDGETOWN	0.0801	2.23	951	393.9	20.76	2.2	12.4
4	CVE 503.0 CANVALE RD	Canvale Rd	CANNING VALE	0.0566	6.82	1,017	11.2	26.9	29.1	17.2
5	CO 506.0 COLLIE BURN	Collie-Burn	COLLIE	0.2596	3.66	569	62.8	11.83	10.5	17.2
6	CC 509.0 JERVOISE BAY	Jervoise Bay	COCKBURN	0.0802	5.76	162	10.9	10.7	4.6	121.7
7	YP 514.0 LACEY RD	Lacey Rd	YANCHEP	0.1869	6.11	2775	55.4	107.2	32.6	7.8
8	MSS 505.0 L327 FREMANTLE RD	L327 Fremantle Rd	MEADOW SPRINGS	0.0539	10.65	4091	44.4	125.7	56	6.5
9	MJP 512.0 DEANMILL	Deanmill	MANJIMUP	0.1254	6.73	1710	270.9	60.38	3.7	12.6



(a)



(b)

Figure 3.6 MV distribution feeder classification on 6 key variables using the modified method, with 9 different symbol for the 9 clusters.

3.5 Summary

This chapter firstly introduced statistical cluster analysis for the distribution feeder classification. The main purpose of it is to determine the optimal number of clusters to be used in the discriminant analysis.

Discriminant analysis was then conducted based upon the result from the cluster analysis. There are differences between the cluster outcomes given by the two statistical analyses suggesting unstable grouping. Combining the two analyses makes it a refined multivariable statistical analysis method, which attempts to obtain a stable classification results with minimum error rate. A set of discriminant functions can be extracted from the method as classifiers

As an illustration for cluster analysis, an example of 25 LV distribution feeders under one representative MV network was shown. The LV networks studied were connected to a metropolitan residential MV feeder within the SWIS in Western Australia. The refined method has then been carried out on a set of 280 MV distribution feeders within SWIS, and a set of discriminant functions has been extracted.

The method introduced in this chapter is flexible and readily transferrable to larger data sets within Western Australia and to other distribution systems that utilize an extensive MV/LV network. For the LV network, it will include many British Commonwealth and European systems that utilize technically similar 230/400V networks. The method has not been evaluated for distribution networks as found in the USA. These 120/240V networks utilize proportionally shorter LV distribution networks and larger numbers of lower rating distribution transformers.

References:

[1] K. P. Schneider, Y. Chen, D. Engle, and D. Chassin, 'A taxonomy of North American radial distribution feeders', in Power & Energy Society General Meeting PES '09, IEEE, *Power & Energy Society General Meeting, 2009. PES '09. IEEE*, 2009, pp. 1-6.

[2] A. Pregelj, M. Begovic, and A. Rohatgi, 'Quantitative techniques for analysis of large data sets in renewable distributed generation', *Power Systems, IEEE Transactions on*, vol. 19, no. 3, pp. 1277-1285, 2004.

[3] G. W. Milligan, *Clustering validation: results and implications for applied analyses*, pp. 341-375, 1996.

[4] G. W. M. a. M. C. Cooper, 'A study of variable standardization', *Journal of Classification*, no. 5, pp. 181-204, 1988.

[5] S. J. Raudys, and A. K. Jain, 'Small sample size effects in statistical pattern recognition: recommendations for practitioners', *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 13, no. 3, pp. 252-264, 1991.

[6] A. El Ouardighi, A. El Akadi, and D. Aboutajdine, 'Feature Selection on Supervised Classification Using Wilks Lambda Statistic', in Computational Intelligence and Intelligent Informatics, 2007. ISCIII '07. International Symposium on, 2007, pp. 51-55.

[7] N. H. Timm, *Applied Multivariate Analysis*, New York: Springer, 2002, pp. 522-523.

[8] K. Schneider, Y. Chen, D. Chassin, R. Pratt, D. Engel, and S. Thompson, 'Modern grid initiative-distribution taxonomy final report', *Pacific Northwest National Laboratory*, 2008.

[9] J. H. Ward, Jr., 'Hierarchical Grouping to Optimize an Objective Function', *Journal of the American Statistical Association*, vol. 58, pp. 236-244, 1963.

[10] M. C. Cooper, and G. W. Milligan, 'The Effect of Error on Determining the Number of Clusters', in Proceedings of the International Workshop on Data Analysis, Decision Support and Expert Knowledge Representation in Marketing and Related Areas of Research, 1988, pp. 319-328.

[11] G. W. Milligan, and M. C. Cooper, 'An Examination of Procedures for Determining the Number of Clusters in a Data Set', *Psychometrika*, vol. 50, pp. 159-179, 1985.

[12] *SAS/STAT® 9.2 User's Guide*, SAS Institute Inc. Cary, NC, USA, 2008.

[13] T. Caliński, and J. Harabasz, 'A dendrite method for cluster analysis', *Communications in Statistics - Theory and Methods*, vol. 3, no. 1, pp. 1-27, 1974.

[14] R. O. Duda, and P. E. Hart, *Pattern classification and scene analysis*, New York: John Wiley & Sons, 1973

[15] L. S. Sheridan J. Coakes, Clara Ong, *SPSS : analysis without anguish : version 16 for Windows*, Australia, 2009.

[16] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Hoboken: John Wiley & Sons, Inc., 2004.

[17] G. De Laurentis, R. Maino, and L. Molteni, *Developing, validating, and using internal ratings: methodologies and case studies*, Chichester, West Sussex, U.K.: Wiley, 2010.

[18] G. A. F. Seber, *Multivariate Observations*, New York: J. Wiley. &. Sone, 1984.

[19] B. F. J. Manly, *Multivariate statistical methods: a primer*: Chapman & Hall /CRC Press, 2005.

[20] G. E. P. Box, 'A general distribution theory for a class of likelihood criteria', *Biometrika*, vol. 36, no. 3-4, pp. 317, 1949.

Chapter 4 Case study of MV/LV distribution feeder classifications

4.1 Introduction

This chapter used the multivariable statistical analysis method introduced in the previous chapter to study two very different data sets within the SWIS – a MV feeder data set of 204 members (22kV) and an LV data set of 8,858 members (230V/400V three phase) that are connected to 204 MV feeders. This efficient statistical method, rather than requiring extensive data sets [1, 2], relies on a few highly meaningful variables that are readily extracted from utility data bases. The available databases [3] from Western Power were reviewed to select the clustering parameters.

4.2 Process of the refined classification method

The case studies include a representative MV feeder set and a representative LV set for an existing network, and presents MV/LV feeder classifiers based upon the refined classification method. The main steps are:

- Data selection –The PNNL approach gathered a large data set with 37 parameters for each feeder [1, 2]. This chapter shows that successful cluster analysis can be economically conducted with far fewer parameters and provides the methods by which the high information value parameters can be determined. The method is applied in two very different test sets – MV and LV feeders.
- Prototypical MV feeder selection – It is shown clustering can be performed with six variables and nine prototypical feeders are identified from the 204 MV (22 kV) feeders under study. The physical feeder closest to its cluster centre is presented as the prototypical MV feeder for the group.

- Prototypical LV feeder selection – After removing inconsistent data, the set under study includes 8,858 physical LV feeders (230V/400V three phase) connected to the 204 MV feeders identified above. Eight prototypical LV feeders are selected using seven key variables. The physical feeder closest to its cluster centre is presented as the prototypical LV feeder.

4.3 Case study of MV feeder classification in WA

4.3.1 MV feeder cluster elements and parameters

The study started with the selection of clustering elements. A data cleaning process first removed feeders where the data sourced from several data bases had any inconsistencies. There are 34 components in the original data base, including:

- Region and substation;
- Peak and mean load (MVA, MVAr, MW) in previous and current year and load growth rate;
- Operating current and voltage;
- Number of customers served;
- Total, MV and LV circuit length(underground, overhead and total);
- Faults, faults per 100 km and customers affected per fault;
- System average interruption duration index, system average interruption frequency index and customer average interruption duration index;
- Total LV distribution transformer number and number of customers per distribution transformer;
- Median LV distribution transformer for each MV feeder (kVA);
- Total installed distribution transformer rating (kVA) along the MV feeder and installed LV distribution transformer capacity per customer.

Selection of variable is important to statistical analysis [4]. Engineering insight suggested an initial set of clustering variables that are thought to be of high importance and a resulting set of clusters determined. The criteria used in cluster parameter selection and dimension reduction were based on engineering knowledge, proximity, contribution, combinability and analysis of variance (ANOVA). For example, regarding the MV feeder clustering, firstly, the parameters such as suburbs and substations were considered less important during clustering and excluded from the cluster parameters. Secondly, some potential clustering variables are highly correlated and only a subset is required for efficient clustering. Peak load and mean load is one such pair. Thirdly, the less sensitive variables were removed. The values of some variables in the study, such as operating voltage, rarely change for different feeders, so their contributions to the cluster were limited and hence excluded from the cluster parameters. The study was restricted to 22kV feeders as this is the dominant feeder voltage and older 6.6kV and 11kV feeders are being systematically retired and replaced by modern 22kV designs. All of the feeders are in the same climatic zone. Fourthly, some variables were combined to one variable to decrease the cluster dimensions. The two variables of the number of customers and the number of distribution transformer were combined into one variable as the new one is more informative. Finally, the efficacy of the selected variables was measured using ANOVA test. Results of ANOVA on MV feeder cluster parameters are listed in Table 4.1. The between groups means were all significant in the ANOVA, indicating each of the six variables reliably distinguished between the six clusters. After several iterations a sufficient set of the significant engineering parameters were clearly identified.

These “high entropy” data variables were identified, from the most significant to the least with regard to clustering, as follows:

- The installed LV distribution transformer capacity per customer, C_{LV} ($F = 126$);
- The MV circuit length, L_{MV} ($F=115$);
- The number of customers per distribution transformer, N_{TX} ($F=86$);
- The number of customers served, N_C ($F=78$);
- The LV circuit length, L_{LV} ($F=52$);
- Peak load, P_L ($F=33$).

TABLE 4.1
ANOVA on MV Feeder Cluster Parameters

		Sum of Squares	**df**	**Mean Square**	***F***	**Sig.**
P_L (MVA)	Between Groups	746.924	8	93.366	33.155	.000
	Within Groups	549.121	195	2.816		
	Total	1296.045	203			
N_C	Between Groups	2.728E8	8	3.410E7	78.624	.000
	Within Groups	8.457E7	195	433671.865		
	Total	3.573E8	203			
L_{MV} (km)	Between Groups	348932.328	8	43616.541	115.453	.000
	Within Groups	73668.351	195	377.786		
	Total	422600.679	203			
L_{LV} (km)	Between Groups	246585.111	8	30823.139	52.178	.000
	Within Groups	115191.954	195	590.728		
	Total	361777.065	203			
N_{TX}	Between Groups	132190.433	8	16523.804	86.215	.000
	Within Groups	37373.460	195	191.659		
	Total	169563.892	203			
C_{LV} (kVA)	Between Groups	105197.077	8	13149.635	126.340	.000
	Within Groups	20295.814	195	104.081		
	Total	125492.891	203			

(df: degree of freedom; Sig.: significance level)

4.3.2 Results of MV feeder cluster analysis

The optimal number of clusters (m_{opt}) needs to be identified by cluster analysis. As mentioned in the previous chapter, the optimal number of clusters is determined taking consideration of *SPRSQ*; Pseudo-*F* Statistic and Pseudo-T^2 Test [5-7].

The favourable number of clusters could be no less than nine according to *SPRSQ* to maintain its large value, as shown in Figure 4.1. High *PSF* values followed by lower values indicate possible cluster solutions, so *PSF* suggests no more than fifteen. When finding a PST^2 value is markedly larger than the previous value, then possible cluster solution could be the one corresponding to the previous value, and the PST^2 recommends three, six, or nine. Therefore, from the above analysis, nine clusters or subsets are selected.

The *x*-axis illustrates the distance at which the clusters merge, while the objects are placed along the *y*-axis as shown in Figure 4.2. The agglomerative hierarchical clustering algorithm were used in the tree graph, starting with 204 clusters, each feeder forms one cluster just as there are 204 ‘branches’ on the ‘tree’, as shown at the left end of the tree graph. The dissimilarities were calculated in order to search for the most similar feeder pair. A new cluster was formed after the pair of feeders combined, which reduced the number of clusters by 1 from 204 to 203. The dissimilarities were calculated again between the newly formed cluster and all remaining clusters until all objects combined into one cluster, as shown at the right end of the tree graph. Cutting the tree graph at a given height offers a clustering at a particular accuracy level. When applying m_{opt} = 9, cutting the tree graph at *SPRSQ* = 0.023 will result in separating the 204 MV feeders into nine subsets. Counting the number of the ‘branches’ at the left end of the tree graph that attached to each cutting point gives the number of feeders in each cluster.

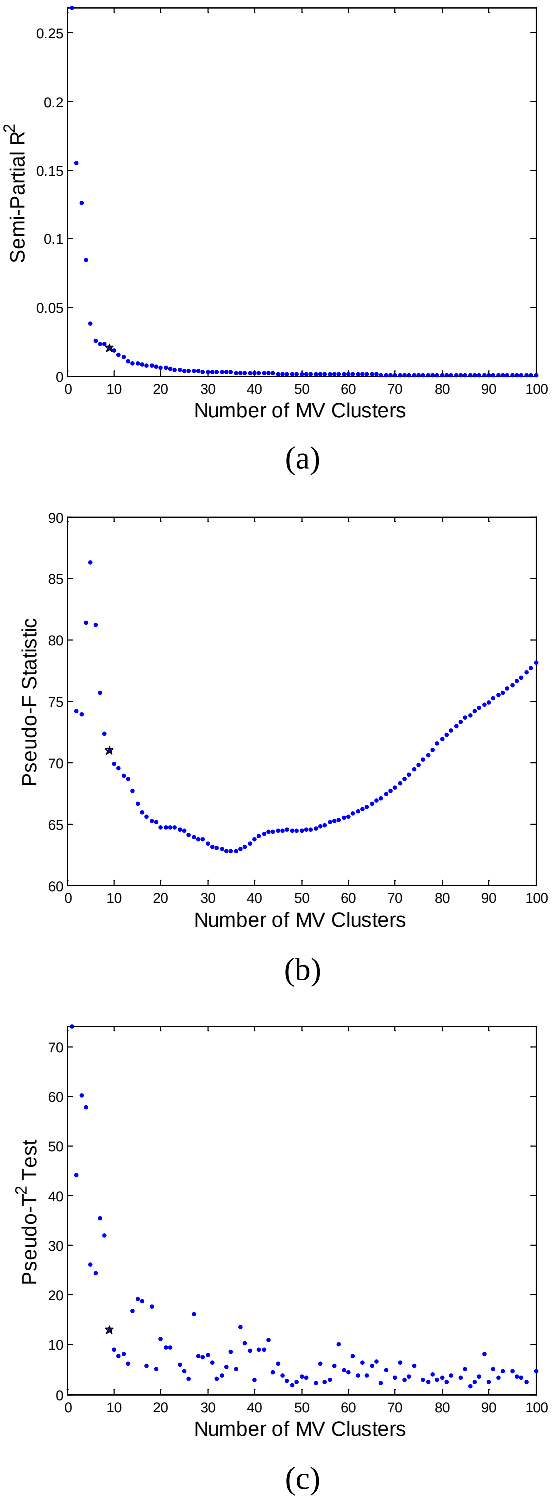


Figure 4.1 Graphs shows the method to determine the optimal MV cluster number, which is 9 (marked with ☆).

(a) – (c) showing the relationship between number of clusters and Semi-Partial R^2, Pseudo-F Statistic and Pseudo-T^2 Test, respectively.

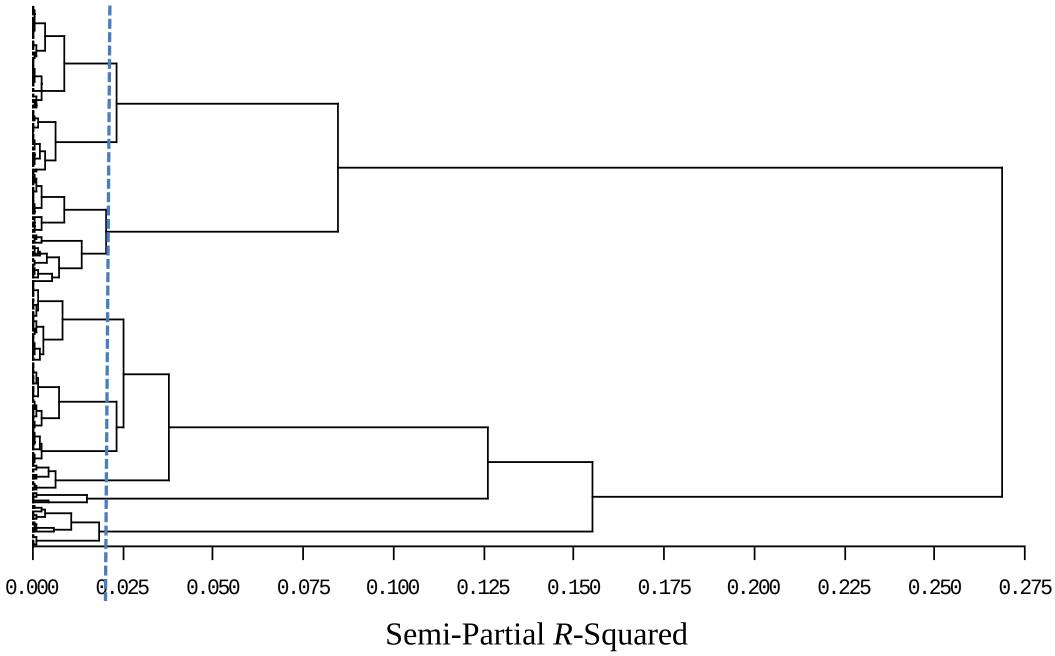


Figure 4.2 Dendrogram resulting from cluster analysis using the Ward's Method for MV feeders.

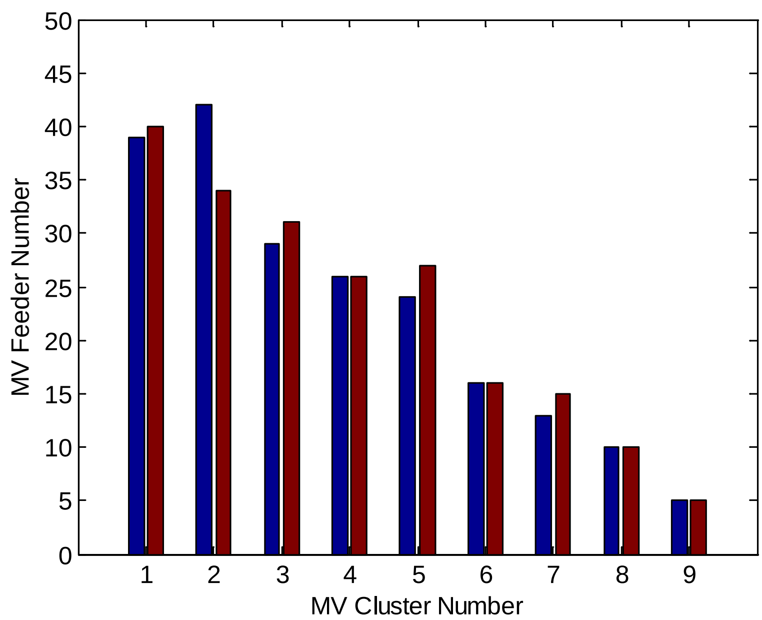


Figure 4.3 A histogram showing the number of MV feeders in each cluster.

The darker histogram (left) shows the number given by cluster analysis, while the lighter histogram (right) shows the one by discriminant analysis.

A histogram displaying the number of feeders in each cluster is shown in Figure 4.3. Two sets of numbers were given, firstly by cluster analysis then by discriminant analysis. As it can be seen, there are some differences between the numbers given by the two statistical analyses.

4.3.3 Results of MV feeder discriminant analysis

a) Pre-tests before building discriminant function

The first step in discriminant analysis is to find out whether the study is valid for DA. If the DFs are found to be statistically significant, the groups can be distinguished based on predictor variables.

(i) Wilks' lambda and multivariate *F* test

Wilks' lambda was used for the significance testing of a set of discriminant functions [7, 8]. The smaller the value, the more it contributes to DF. As seen in Table 4.2, the level of significance is below 0.005, indicating the six feeder variables are significantly different and will contribute strongly to the construction of DFs. The assumption of equality of covariances across groups does not apply and the DFs should be quadratic rather than linear.

TABLE 4.2
Tests of Equality of Group Means

	Λ	***F***	**df1**	**df2**	**Sig.**
P_L	0.424	33.2	8	195	0.000
N_C	0.237	78.6	8	195	0.000
L_{MV}	0.174	115.5	8	195	0.000
L_{LV}	0.318	52.2	8	195	0.000
N_{TX}	0.220	86.2	8	195	0.000
C_{LV}	0.162	126.2	8	195	0.000

(df: degree of freedom; Sig.: significance level; *F*: a test to identify which variable will most separate the group)

(ii) Tests for homogeneity of covariance matrices

Tests for homogeneity of covariance matrices are essential in order to assess whether linear or quadratic DFs are applicable in this case. As seen in Table 4.3, the level of significance is below 0.005. The assumption of equality of covariances across groups does not apply and the DFs should be quadratic rather than linear.

TABLE 4.3
Box's *M* Test Results

Box's *M*		1540.2
F	Approx.	9.006
	df1	147
	df2	12702
	Sig.	0.000
Tests null hypothesis of equal population covariance matrices.		

b) Building DFs

DF can be used as a feeder classifier [9]. The initial clustering analysis is the starting point for the DA process. Hierarchical single-pass clustering cannot correct for erroneous decisions made in a first pass. After the initial DA, the classification results have some differences to those found in the primary cluster analysis, indicating the first cluster had unstable features. Iterative rounds of DA and DF construction were applied until stable cataloguing was achieved at the seventh iteration. The DF extracted in the last round can be used as a classifier for any MV feeder in the distribution network. As mentioned in the previous chapter, the QDF can be expressed in the form in Eq. (3.19), where, in this case, $k = 1,\ldots,m_{opt}$ ($m_{opt} = 9$); and $d = 6$. x_i and x_j (i, j = 1,…,6) are the six key variables. The coefficients of a_0, a_i and a_{ij} for each cluster are listed in Table A.2 in Appendix. Nine calculated values (i.e. $G_1,\ldots,G_9$) are determined for each MV feeder using Eq. (3.19). The 204 feeders are classified into nine clusters according to the highest classification scores G_k. For example, using Eq. (3.19), the QDF of Cluster 1 can be written as:

$$
\begin{aligned}
G(x) = {} & a_0 + a_1x_1 + a_{11}x_1x_1 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4 + a_{15}x_1x_5 + a_{16}x_1x_6 \\
& + a_2x_2 + a_{22}x_2x_2 + a_{23}x_2x_3 + a_{24}x_2x_4 + a_{25}x_2x_5 + a_{26}x_2x_6 \\
& + a_3x_3 + a_{33}x_3x_3 + a_{34}x_3x_4 + a_{35}x_3x_5 + a_{36}x_3x_6 \\
& + a_4x_4 + a_{44}x_4x_4 + a_{45}x_4x_5 + a_{46}x_4x_6 \\
& + a_5x_5 + a_{55}x_5x_5 + a_{56}x_5x_6 \\
& + a_6x_6 + a_{66}x_6x_6
\end{aligned}
\tag{4.1}
$$

4.3.4 Analysis of MV subsets and representative MV feeders

The representative MV feeder in a specific cluster is the feeder closest to its cluster centre. Representative feeders were identified through normalising data by *Z*-score transformation followed by comparing *SEUCID*. The summary of the nine clusters is listed in Table 4.4, showing the mean value for the six key parameters in each cluster. The nine clusters are designated with varying geometrical symbols in Figure 4.4, which gives a high separation of the clusters in the three dimensional view. Figure 4.5 displays the physical locations of the nine MV prototypical feeders. Descriptions of prototypical MV Feeders are listed in Table 4.5, and the characteristics of the representative MV feeders are in Table 4.6. Clusters 1, 2, 4, 5 and 7 are residential feeders. Cluster 3 is best described as mixed metropolitan feeder. Cluster 6 is a commercial and/or industrial feeder. Cluster 8 and 9 are rural feeders. The details are as the follows.

Five clusters are residential feeders:

- 1 – Light urban residential 1 – 2,719 customers, 5.40kVA per customer, 9.45MVA peak load, representative feeder H 517.0 PALMERSTON ST;
- 2 – Heavy urban residential – 3,983 customers, 5.95kVA per customer, 10.18MVA peak load, representative feeder MSS 505.0 L327 FREMANTLE RD;
- 4 – Moderate suburban residential – 2,299 Customers, 6.13kVA per customer, 6.56MVA peak load, representative feeder WE 519.0 1 FELSPAR RD;
- 5 – Light urban residential 2 – 3,179 customers, 4.31kVA per customer, 8.13MVA peak load representative feeder WNO 506.0 L306 WANNEROO RD;

- 7 – Light suburban residential – 1,565 customers, 5.07kVA per customer, 4.98MVA peak load representative feeder OC 514.0 ANTILL ST;

- One feeder is best described as mixed metropolitan feeder:

- 3 – Heavy mixed metropolitan – 1,049 customers, 14.79kVA per customer, 7.20MVA peak load, representative feeder CVE 503.0 CANVALE RD;

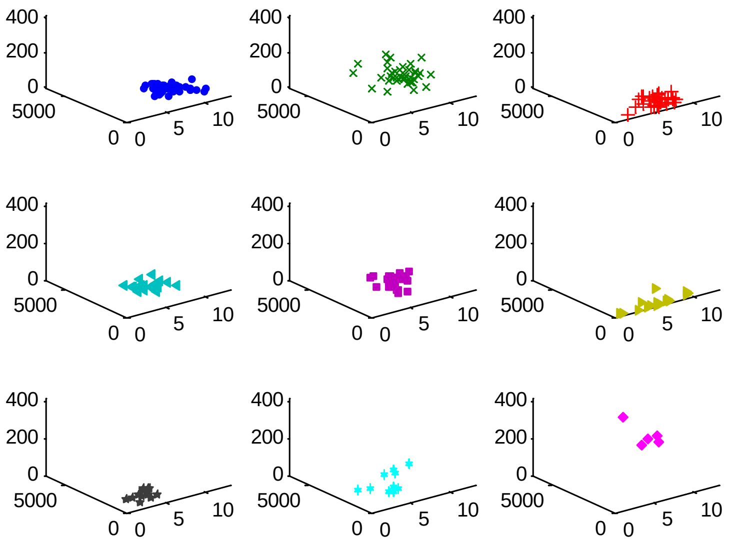
One feeder is a commercial and/or industrial feeder:

- 6 – Commercial/industrial – 185 customers, 91.01kVA per customer, 5.76MVA peak load, representative feeder H 504.0 COLLIER RD;

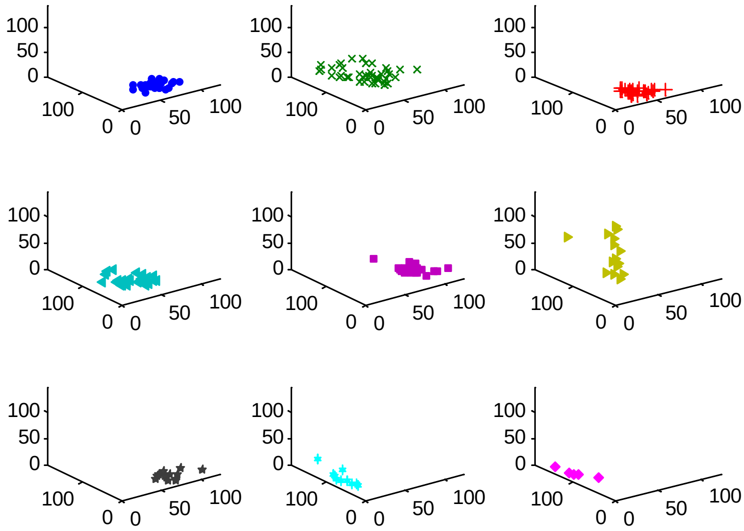
The last two clusters are rural feeders:

- 8 – Rural short – 1,226 customers, 12.49kVA per customer, 4.17MVA peak load, representative feeder K 502.0 LEWIS RD EAST;

- 9 – Rural long – 2,040 customers, 14.32kVA per customer, 7.22MVA peak load, MV Length 270km, representative feeder SV 501.0 CHIDLOW.



(a) 3D plot of P_L(MVA)(X), N_c(Y) and L_{MV}(km)(Z)



(b) 3D plot of $N_{TX}(X)$, L_{LV}(km)(Y) and C_{LV}(kVA)(Z)

Figure 4.4 MV feeder clustering on 6 key variables, with 9 subplots for the 9 clusters or subsets.

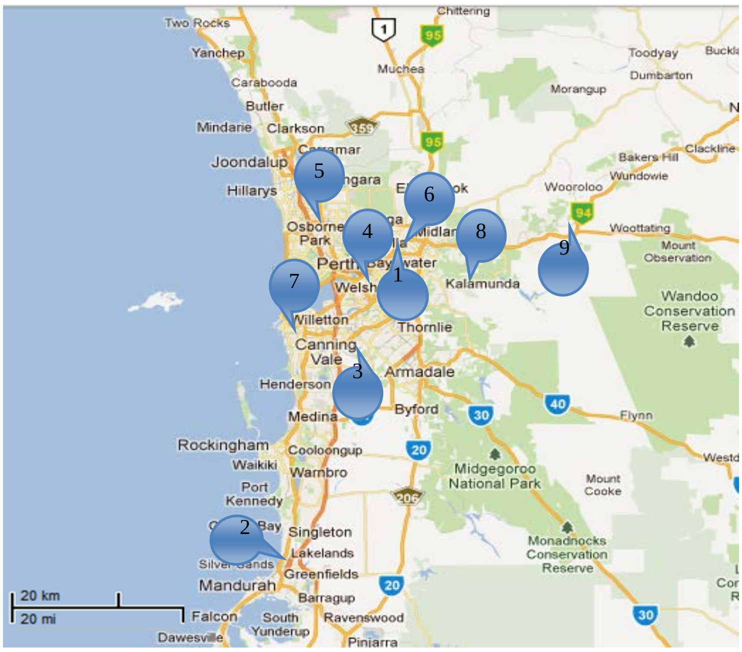


Figure 4.5 Physical locations of the nine MV prototypical feeders

TABLE 4.4
Summary of MV Feeders Classification after Discriminant Analysis

Cluster	1	2	3	4	5	6	7	8	9
P_L (MVA)	9.45	10.18	7.20	6.56	8.13	5.76	4.98	4.17	7.22
N_C	2719	3983	1049	2299	3179	185	1565	1226	2040
L_{MV} (km)	18.70	41.23	11.34	27.79	17.45	12.35	10.37	80.85	270.4
L_{LV} (km)	55.81	128.8	24.21	63.35	73.35	18.32	33.64	79.69	102.1
N_{TX}	64.81	57.86	37.87	46.47	94.70	6.39	78.28	10.64	4.88
C_{LV} (kVA)	5.40	5.95	14.79	6.13	4.31	91.01	5.07	12.49	14.32
No. of Feeders	40	34	31	26	27	16	15	10	5

TABLE 4.5
Descriptions of MV Feeder Subsets

Cluster	Description
1	Light urban residential 1
2	Heavy urban residential
3	Heavy mixed metropolitan
4	Moderate suburban residential
5	Light urban residential 2
6	Commercial/Industrial
7	Light suburban residential
8	Rural short
9	Rural long

TABLE 4.6
Characteristics of the Nine Representative MV Feeders

Cluster	Feeder	Street	Substation	SEUCID	P_L (MVA)	N_C	L_{MV} (km)	L_{LV} (km)	N_{TX}	C_{LV} (kVA)
1	H 517.0 PALMERSTON ST	Palmerston St	HADFIELDS	1.022	9.64	2,673	18.9	55.4	54.6	6.1
2	MSS 505.0 L327 FREMANTLE RD	L327 Fremantle Rd	MEADOW SPRINGS	0.404	10.65	4,091	44.4	125.7	56.0	6.5
3	CVE 503.0 CANVALE RD	Canvale Rd	CANNING VALE	.841	6.82	1,017	11.2	26.9	29.1	17.2
4	WE 519.0 1 FELSPAR RD	1 Felspar Rd	WELSHPOOL	.709	6.68	2,117	26.0	72.3	44.1	6.1
5	WNO 506.0 L306 WANNEROO RD	L306 Wanneroo Rd	WANNEROO	.875	8.71	3,033	18.5	77.5	89.2	4.9
6	H 504.0 COLLIER RD	Collier Rd	HADFIELDS	.932	7.17	234	7.2	13.1	7.8	80.0
7	OC 514.0 ANTILL ST	Antill	OCONNOR	1.021	5.45	1,664	7.8	31.7	72.3	5.6
8	K 502.0 LEWIS RD EAST	Lewis East	KALAMUNDA	1.373	4.27	908	59.8	51.1	9.1	10.6
9	SV 501.0 CHIDLOW	Chidlow	SAWYERS VALLEY	1.020	9.85	2,790	221.10	151.74	6	7.2

4.4 LV distribution feeder classification

In many systems, notably those outside of the USA, the LV network may be physically extensive. The LV network structure and impedance will often have a larger impact on consumer power quality than the MV network. In these systems, the proliferation of distributed generation resources within the LV grids, especially domestic photovoltaic systems, has been modifying the traditional network operating parameters [10]. Some LV feeders service dozens of customers, yet some others serve hundreds or even thousands of customers. Figure 4.6 shows the number of customers that serviced by the LV feeders in Western Australia. Considering the complexity of the smart grid networks, efficient modeling and simulation are crucial [11]. As most of the voltage regulation occurs at LV, it could be possible to extract common LV constructions, and perform cluster analysis.

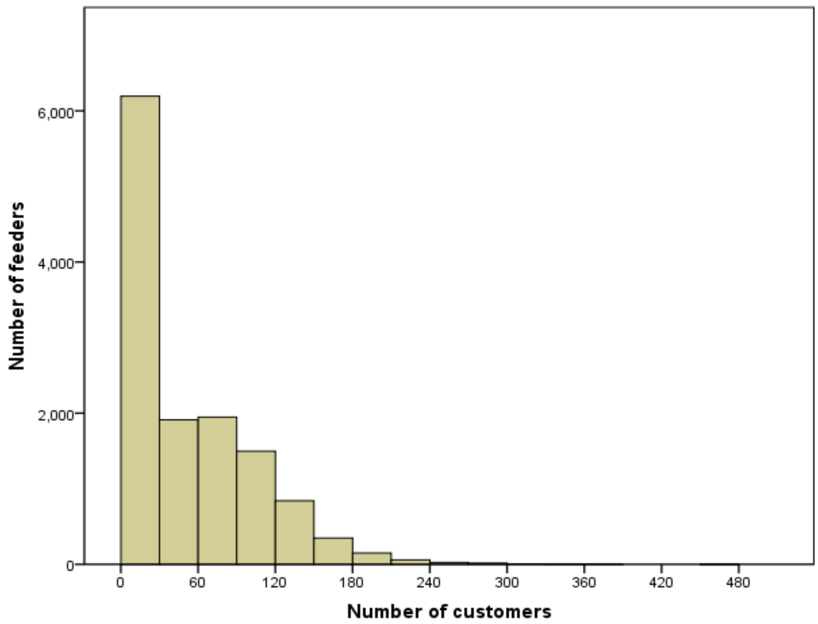


Figure 4.6 A histogram showing the number of customers serviced by the LV feeders in Western Australia.

4.4.1 LV feeder cluster elements and parameters

The feeder set includes 8,858 physical LV feeders (230V/400V three phase) connected to 204 MV feeders. The available database [3] from Western Power was reviewed to select the clustering parameters. There are 26 components in the original database, including:

- LV transformer rating;
- Feeder lengths;
- Energy consumptions in each class (residential, commercial, industrial and 24-hour-users, respectively);
- Total number of days of customer connection (for each category users);
- Customer numbers (for each category of users);
- Load capacity based on load factor and energy consumption (from residential, commercial, industrial and 24-hour-user, respectively and in total);
- Calculated residential ratio based on consumptions;
- Conservative residential and transformer load capacity calculation;
- ADMD;
- Residential load ratio;
- Total load capacity/Rating ratio.

Not all of the components were included in the cluster analysis. Same as the MV feeder case, the criteria used in dimension reduction are based on engineering knowledge, proximity, contribution, combinability and ANOVA test.

Take the parameters of residential energy consumption and load capacity as an example. Only one of them was selected after the Spearsman Correlation test (correlation coefficient = 0.997), as the two variables were tested to be correlated. The results of the Spearsman Correlation test are shown in Table 4.7. Residential load ratio is a less sensitive parameter and was excluded. Energy consumptions in classes of commercial, industrial and 24-hour-users were combined to one variable. Similarly, load capacities and the number of customers in the three classed were combined as well.

Results of ANOVA on MV feeder cluster parameters are listed in Table 4.8, showing the between groups means were all significant in the ANOVA, and each of the seven feeder parameter reliably distinguished between the eight clusters.

The informative variables selected, from the strongest differentiator to the lowest, are:

- Underground feeder length, L_{UG} (F=4755);
- Load capacity of residential customers, P_r (F=4732);
- Number of residential customers, N_r (F=3857);
- Overhead feeder length, L_{OH} (F=2552);
- Total load capacity/Rating ratio, R (F=2264);
- Load capacity excluding residential customers, P_{others} (F=1937);
- Number of customers excluding residential customers, N_{others} which includes number of commercial, industrial and 24/7-hour customers (F=1853).

TABLE 4.7
Results of Spearsman Correlation Test

		Residential energy consumption	Residential load capacity
Residential energy consumption	Correlation Coefficient	1.000	.997**
	Sig. (2-tailed)	.	.000
	N	8858	8858
Residential load capacity	Correlation Coefficient	.997**	1.000
	Sig. (2-tailed)	.000	.
	N	8858	8858

**. Correlation is significant at the 0.01 level (2-tailed).

TABLE 4.8
ANOVA on LV Feeder Cluster Parameters

		Sum of Squares	df	Mean Square	F	Sig.
	Between Groups	322034.524	7	46004.932	1853.891	.000
N_{others}	Within Groups	219615.751	8850	24.815		
	Total	541650.275	8857			
	Between Groups	1.584E7	7	2263500.580	3856.635	.000
N_r	Within Groups	5194160.727	8850	586.911		
	Total	2.104E7	8857			
	Between Groups	3.210E7	7	4585544.516	1937.997	.000
P_{others} (kVA)	Within Groups	2.094E7	8850	2366.126		
	Total	5.304E7	8857			
	Between Groups	1.061E8	7	1.515E7	4732.983	.000
P_r (kVA)	Within Groups	2.833E7	8850	3201.296		
	Total	1.344E8	8857			
	Between Groups	838.605	7	119.801	2263.755	.000
R (%)	Within Groups	468.353	8850	.053		
	Total	1306.958	8857			
	Between Groups	1.735E9	7	2.478E8	2551.711	.000
L_{OH} (m)	Within Groups	8.595E8	8850	97123.386		
	Total	2.594E9	8857			
	Between Groups	9.053E9	7	1.293E9	4755.518	.000
L_{UG} (m)	Within Groups	2.407E9	8850	271966.858		
	Total	1.146E10	8857			

The load capacity/rating ratio *R* is an estimate of the distribution transformer loading at the peak time. Figure 4.7 shows the number of LV feeders that each of the 204 MV feeder supplies. It can be seen that some MV feeders supply dozens of LV feeders, while some others supply hundreds of LV feeders.

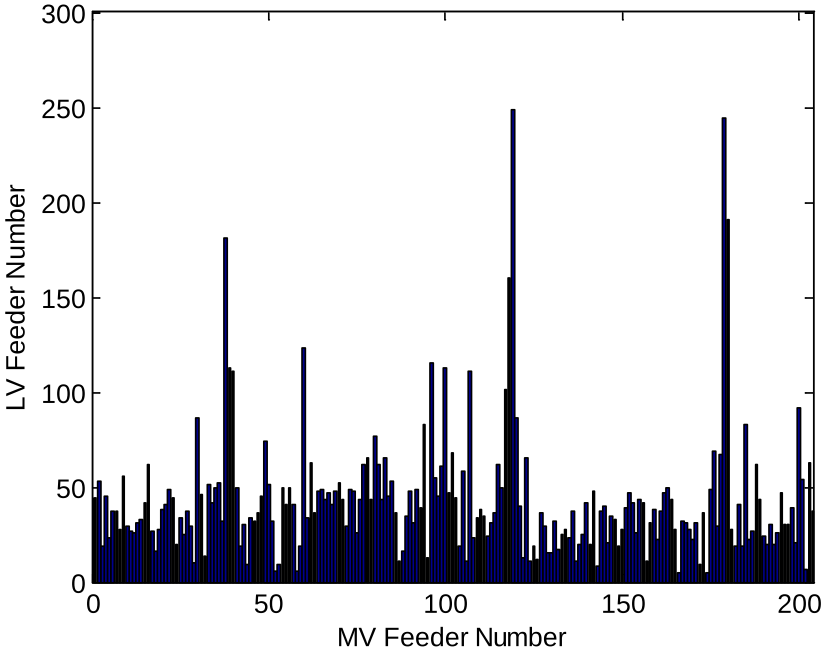
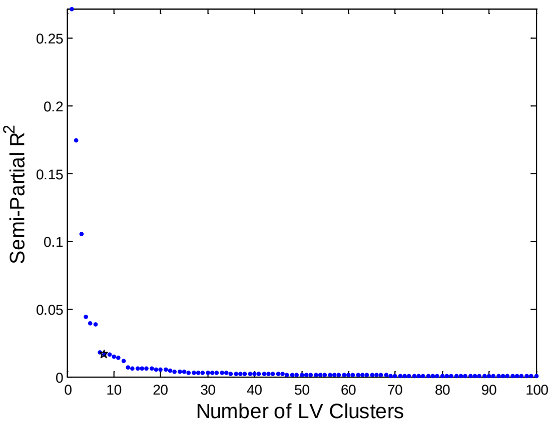
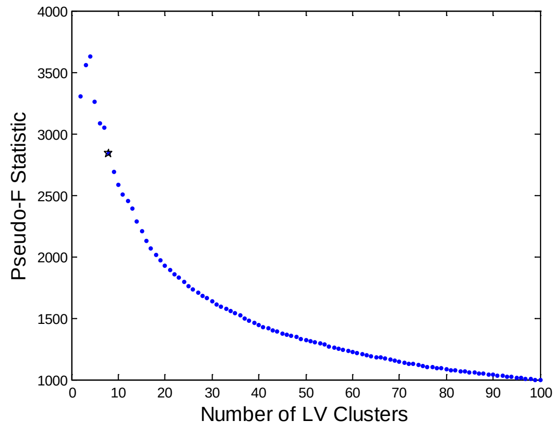


Figure 4.7 A histogram showing the number of LV feeders that each of the 204 MV feeder supplies.

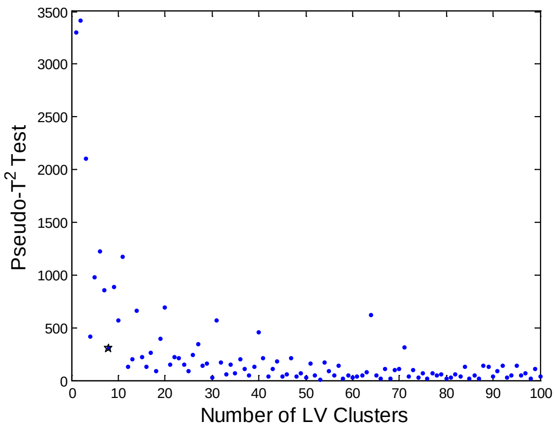
4.4.2 Results of LV feeder cluster analysis



(a)



(b)



(c)

Figure 4.8 Graphs shows the method to determine optimal LV cluster number.

The optimal number of clusters (m_{opt}) is determined to be eight (marked with ☆), when taking consideration of the three parameters of Semi-Partial R^2, Pseudo-F Statistic and Pseudo-T^2 Test.

As shown in Figure 4.8, the favourable number of clusters could be no less than seven according to *SPRSQ*. The pseudo-F suggests no more than twenty, and the pseudo-T^2 recommends four, eight, or twelve. Therefore, from the above analysis, eight clusters or subsets are selected.

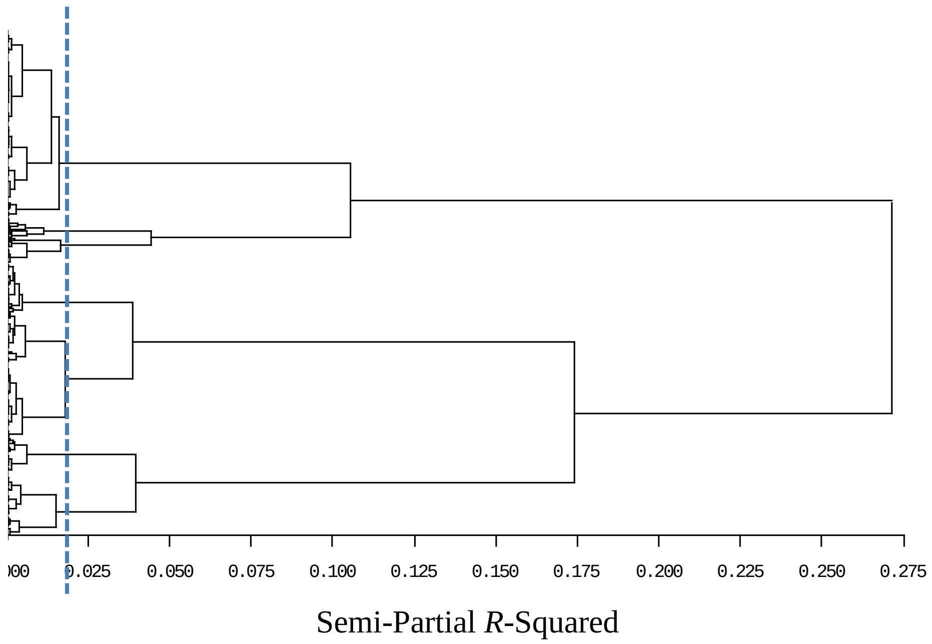


Figure 4.9 Dendrogram resulting from cluster analysis using the Ward's Method for LV distribution feeders.

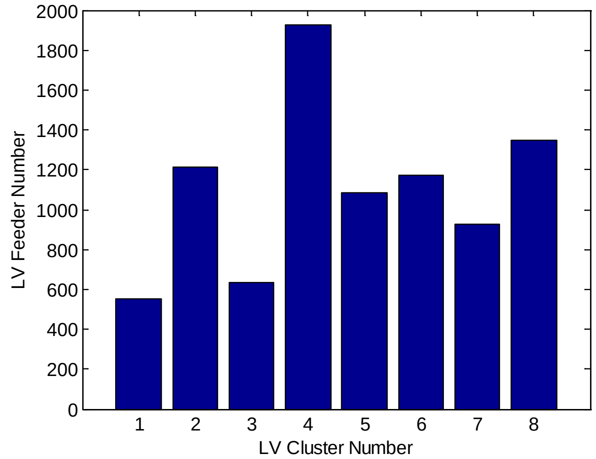


Figure 4.10 Histogram showing the number of LV feeders in each cluster determined by discriminant analysis.

The LV optimal cluster number is eight when the tree graph being cut at *SPRSQ* = 0.017 as shown in Figure 4.9. A histogram displaying the number of LV feeders in each cluster is shown in Figure 4.10.

4.4.3 Results of LV feeder discriminant analysis

a) Pre-tests before building discriminant function

Similar to the MV case study, it is essential to find out whether the study is valid for DA before building the DFs.

(i) Wilks' lambda and multivariate *F* test

As seen in Table 4.9, the significance level of each variable is under 0.005, indicating the six feeder variables are significantly different and will contribute strongly to the construction of DFs.

TABLE 4.9
Tests of Equality of Group Means

	Λ	***F***	**df1**	**df2**	**Sig.**
N_{others}	0.614	795.6	7	8850	0.000
N_r	0.633	732.5	7	8850	0.000
P_{others} (kVA)	0.643	700.8	7	8850	0.000
P_r (kVA)	0.623	765.6	7	8850	0.000
R (%)	0.709	519.2	7	8850	0.000
L_{OH} (m)	0.409	1829.3	7	8850	0.000
L_{UG} (m)	0.464	1460.2	7	8850	0.000

(df: degree of freedom; Sig.: significance level; *F*: a test to identify which variable will most separate the group)

(ii) Tests for homogeneity of covariance matrices

As seen in Table 4.10, the level of significance is below 0.005. The assumption of equality of covariances across groups does not apply and the DFs should be quadratic rather than linear.

TABLE 4.10
Box's *M* Test Results

Box's *M*		46795.8
F	Approx.	416.3
	df1	112
	df2	2.108E7
	Sig.	0.000
Tests null hypothesis of equal population covariance matrices.		

b) Building DFs

Regarding the QDF, the coefficients of a_0, a_i and a_{ij} in Eq. (3.19) for each cluster are listed in Table A.3 in Appendix. Eight calculated values (i.e. $G_1,\ldots,G_8$), are determined for each LV feeder using Eq. (3.19). The 8,858 feeders are classified into eight clusters according to the highest classification scores G_k. For example, the QDF of Cluster 1 can be written as:

$$
\begin{aligned}
G(x)_{k=1} = {} & a_0 + a_1x_1 + a_{11}x_1x_1 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4 + a_{15}x_1x_5 + a_{16}x_1x_6 + a_{17}x_1x_7 \\
& + a_2x_2 + a_{22}x_2x_2 + a_{23}x_2x_3 + a_{24}x_2x_4 + a_{25}x_2x_5 + a_{26}x_2x_6 + a_{27}x_2x_7 \\
& + a_3x_3 + a_{33}x_3x_3 + a_{34}x_3x_4 + a_{35}x_3x_5 + a_{36}x_3x_6 + a_{37}x_3x_7 \\
& + a_4x_4 + a_{44}x_4x_4 + a_{45}x_4x_5 + a_{46}x_4x_6 + a_{47}x_4x_7 \\
& + a_5x_5 + a_{55}x_5x_5 + a_{56}x_5x_6 + a_{57}x_5x_7 \\
& + a_6x_6 + a_{66}x_6x_6 + a_{67}x_6x_7 \\
& + a_7x_7 + a_{77}x_7x_7
\end{aligned}
\tag{4.2}
$$

4.4.4 Analysis of LV subsets and representative LV feeders

The summary of the eight clusters is listed in Table 4.11, showing the mean value for the seven key parameters in each cluster.

TABLE 4.11
Summary of LV Feeders Classification after discriminant analysis

Cluster	1	2	3	4	5	6	7	8
N_{others}	0.80	2.04	3.04	0.49	15.53	5.85	0.22	0.50
N_r	2.20	59.55	100.9	61.68	28.73	0.00	22.73	69.36
P_{others} (kVA)	3.88	8.90	23.86	0.50	117.9	101.14	0.20	0.37
P_r (kVA)	7.36	135.9	264.5	170.7	67.55	0.00	62.79	162.7
R (%)	25.34	71.27	68.83	58.16	48.50	20.41	79.41	81.55
L_{OH} (m)	208.4	1011	244.9	0.00	516.8	82.69	675.4	1058
L_{UG} (m)	5.38	320.3	2391	1893	477.1	171.8	52.69	166.9
Number of Feeders	550	1212	637	1925	1082	1174	928	1350

Characteristics of the representative LV feeders are listed in Table 4.12, and the descriptions of prototypical LV Feeders are given in Table 4.13. The eight clusters are designated with varying geometrical symbols in Figure 4.11, which gives a high separation of the clusters in the three dimensional view. Physical locations of the eight LV prototypical feeders are displayed in Figure 4.12.

Just as shown in the descriptions of the eight prototypical LV feeders in Table 4.15, Clusters 1, 5 and 6 are feeders servicing a mix of customer types; while Clusters 2, 3 4, 7 and 8 are residential feeders. The details are as the follows.

The following five clusters are residential feeders:

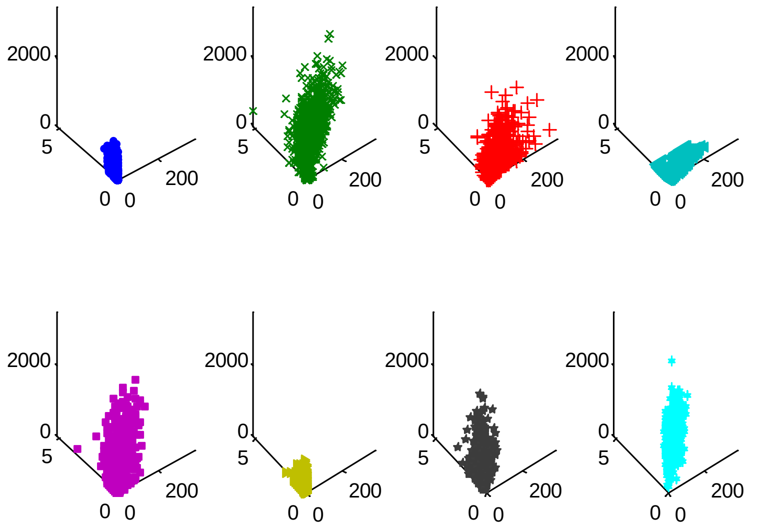
- 2 –Residential with mainly using overhead lines, mainly using underground lines –1,011m overhead feeder length, 59.55 residential customers, 135.9kVA residential peak load, 2.04 others customers, 8.90kVA others customer types' peak load, average 71.27% transformer utilisation rate. There are 2,644 LV feeders in this group, and the prototypical feeder is 12 JANNALI WAY ;
- 3 –Residential with high transformer utilisation, mainly using underground lines– 2,391m underground feeder length, 100.9 residential customers, 264.5kVA residential peak load, 3.04 others customers, 23.86kVA others customer types' peak load, average 68.83% transformer utilisation rate. There are 1,450 LV feeders in this group, and the prototypical feeder is CRANBERRY;

TABLE 4.12
Characteristics of the Eight Representative LV Feeders

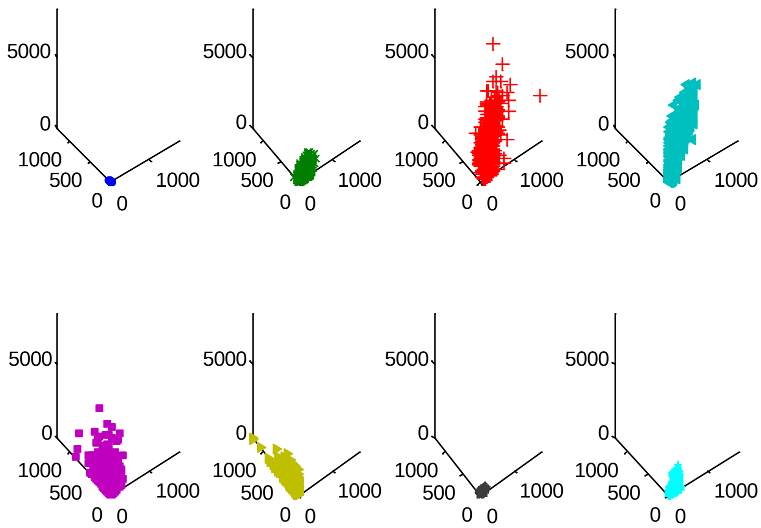
Cluster	**MV Feeder Name**	**LV Feeder Name**	N_{others}	N_r	P_{others} *(kVA)*	P_r *(kVA)*	*R (%)*	L_{OH} *(m)*	L_{UG} *(m)*	*SEUCID*
1	BYF 523.0 L4085 SUNRAYS ST	CLARKE	1	2	2.26	5.00	29.03	189.5	0.00	0.76
2	MSS 508.0 MEADOW SPRINGS S/S	12 JANNALI WAY	2	59	4.93	137.1	71.01	1194	335.4	0.49
3	WE 504.0 CANNINGTON	CRANBERRY	4	101	7.02	239.1	78.11	214.4	1985	0.74
4	NB 519.0 BEACH RD WEST	FREEMAN 2	1	71	0.33	178.3	56.70	0.00	1842	0.70
5	WE 523.0 WELSHPOOL WEST	WELSHPOOL 5	18	19	69.11	84.07	51.06	582.3	457.2	0.62
6	KDL 505.0 MACKAY ST	KINGSCOTE	5	0	68.36	0.00	13.67	100.1	100.9	0.49
7	K 504.0 GUPPY RD	CRUMPET CREEK	0	19	0.00	64.06	64.06	622.9	39.65	0.82
8	CC 501.0 RUSSELL RD WEST	KIESEY	1	76	0.12	154.7	77.42	1058	190.3	0.81

TABLE 4.13
Descriptions of LV Feeder Subsets

Cluster	**Description**
1	Small commercial/residential with low transformer utilisation and short lines
2	Residential with mainly using overhead lines
3	Residential with high transformer utilisation, mainly using underground lines
4	Residential with low transformer utilisation, only using underground lines
5	Median mixed commercial/residential with mainly using overhead lines
6	Medium commercial/industrial with short LV lines
7	Small Residential with high transformer utilisation, using short total LV lines
8	Residential with high transformer utilisation mainly using overhead lines



(a) 3D plot of Nr(X), R (%)(Y) and LOH(m)(Z)



(b) 3D plot of P_r(kVA)(X), P_{others}(kVA)(Y) and L_{UG}(m)(Z)

Figure 4.11 The 8,858 LV feeders clustering on 7 key variables, with 8 subplots for the 8 clusters. (only 6 variables showing in Figure (a) and (b))

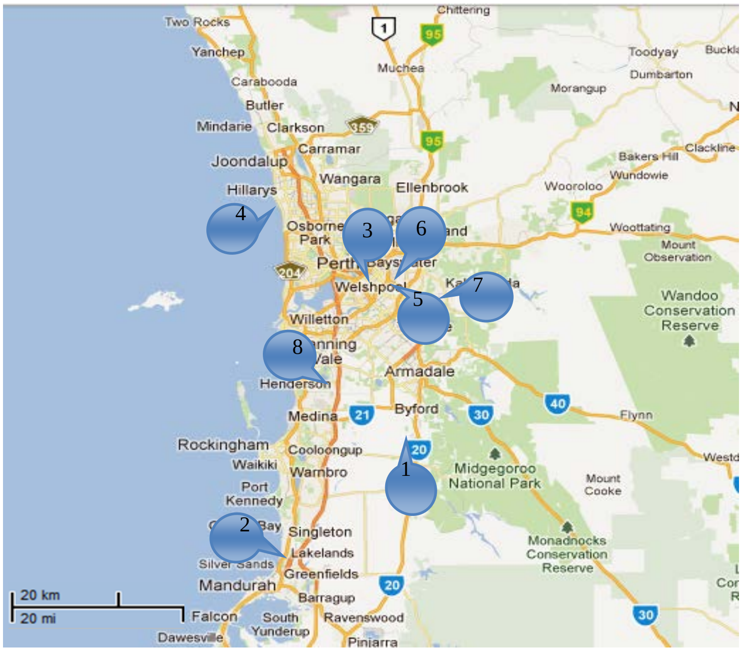


Figure 4.12 Physical locations of the eight LV prototypical feeders

- 4 –Residential with low transformer utilisation, only using underground lines – 1,893m overhead feeder length, 61.68 residential customers, 170.7kVA residential peak load, 0.49 others customers, 0.50kVA others customer types' peak load, average 58.16% transformer utilisation rate. There are 1,925 LV feeders in this group, and the prototypical feeder is FREEMAN 2;
- 7 –Small Residential with high transformer utilisation, using short total LV lines– only 728m total feeder length, 22.73 residential customers, 62.79kVA residential peak load, average 79.41% transformer utilisation rate. There are 928 LV feeders in this group, and the prototypical feeder is CRUMPET CREEK;
- 8 –Residential with high transformer utilisation mainly using overhead lines – 1,058m overhead feeder length, 166.9m underground feeder length, 69.4 residential customers, 0.50 others customers, 162.7kVA residential peak load, 0.37kVA others customer types' peak load, average 81.55% transformer

utilisation rate. There are 1,350 LV feeders in this group, and the prototypical feeder is KIESEY;

The following three clusters include feeders servicing a mix of customer types:

- 1 –Small commercial/residential with low transformer utilisation and short lines – 208.4m overhead feeder length, 5.38m underground feeder length, 2.2 residential customers, 0.8 others customers, 7.36kVA residential peak load, 3.88kVA others customer types' peak load, average 25.34% transformer utilisation rate. There are 550 LV feeders in this group, and the prototypical feeder is CLARKE;
- 5 –Median mixed commercial/residential with mainly using overhead lines – 516.8m overhead feeder length, 477.0m underground feeder length, 28.73 residential customers, 15.53 others customers, 67.55kVA residential peak load, 117.9kVA others customer types' peak load, average 48.50% transformer utilisation rate. There are 1,082 LV feeders in this group, and the prototypical feeder is WELSHPOOL 5 ;
- 6 –Medium commercial/industrial with short LV lines –82.69m overhead feeder length, 171.8m underground feeder length, 0 residential customers, 5.85 others customers, 0kVA residential peak load, 101.1 others customer types' peak load, average 20.41% transformer utilisation rate. There are 1,174 LV feeders in this group, and the representative feeder KINGSCOTE.

4.5 Summary

This chapter used statistical methods introduced in the previous chapter to identify statistically significant feeders with a limited number of engineering variables. Distribution utilities maintain data bases that can provide relevant clustering information. The iterative methods can be applied to rigorously determine which variables are valuable from a clustering perspective. Once a robust set of input variables is determined a combination of Ward's hierarchical clustering method and discriminant analysis can produce stable clusters and feeder classifiers.

The methods are illustrated using two very different data bases – a MV feeder data base of 204 members and an LV data base of 8,858 members. The capability of the general method to form useful cluster classifications in such disparate applications illustrates the universality of the methods. This is a highly transferrable approach that will be applicable in many distribution networks.

The nine prototypical MV distribution feeders made use of the six selected variables. A set of discriminant functions has been extracted that can be used as the classifier for new MV feeders to be built in the WA distribution networks. Similarly, for the LV distribution feeders, eight prototypical LV feeders have been identified, making use of seven variables. The WA distribution network under study is within one climate zone, and weather related parameters play less important role in our case studies, hence these parameters (such as temperature) are not incorporated in the clustering process. However, climate or weather related parameters may be needed for case studies in other countries.

Once a statistically significant set of feeders is determined for any given network, it is possible to use these to efficiently evaluate a range of smart grid technologies that could be applied. Two recent examples are the application of the PNNL test feeder set to provide a rapid, and statistically defensible, evaluation of Conservation Voltage Reduction (CVR), across the US distribution network [12] and the use of the UKGDS models to evaluate demand management and energy storage [13]. Finally we'd like to progressively make the representative feeder models freely available on-line through the Power and Energy Centre at CQU. We intend to provide these in a node and feeder form following the IEEE test set format.

References:

[1] K. P. Schneider, Y. Chen, D. Engle, and D. Chassin, 'A taxonomy of North American radial distribution feeders', in Power & Energy Society General Meeting PES '09, IEEE, *Power & Energy Society General Meeting, 2009. PES '09. IEEE*, 2009, pp. 1-6.

[2] K. P. Schneider, Y. Chen, D. P. Chassin, R. G. Pratt, D. W. Engel, and S. E. Thompson, *Modern grid initiative-distribution taxonomy final report*, Pacific Northwest National Laboratory, 2008.

[3] *Metro distribution transformer load forecast report*, Western Power Corporation, WA, 2010.

[4] S. J. Raudys, and A. K. Jain, 'Small sample size effects in statistical pattern recognition: recommendations for practitioners', *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 13, no. 3, pp. 252-264, 1991.

[5] G. W. Milligan, and M. C. Cooper, 'An Examination of Procedures for Determining the Number of Clusters in a Data Set', *Psychometrika*, vol. 50, pp. 159-179, 1985.

[6] M. C. Cooper, and G. W. Milligan, 'The Effect of Error on Determining the Number of Clusters', in Proceedings of the International Workshop on Data Analysis, Decision Support and Expert Knowledge Representation in Marketing and Related Areas of Research, 1988, pp. 319-328.

[7] *SAS/STAT® 9.2 User's Guide*, SAS Institute Inc. Cary, NC, USA, 2008.

[8] L. S. Sheridan J. Coakes, Clara Ong, *SPSS : analysis without anguish : version 16 for Windows*, Australia, 2009.

[9] R. O. Duda, and P. E. Hart, *Pattern classification and scene analysis*, New York: John Wiley & Sons, 1973

[10] S. Papathanassiou, N. Hatziargyriou, and K. Strunz, 'A Benchmark Low Voltage Microgrid Network', in CIGRE Symposium, Athens, 2005.

[11] Y. Li, and P. Wolfs, 'Statistical Identification of Prototypical Low Voltage Distribution Feeders in Western Australia', in Power and Energy Society General Meeting, IEEE, 2012, pp. 1-8.

[12] K. P. Schneider, J. C. Fuller, F. K. Tuffner, and R. Singh, *Evaluation of Conservation Voltage Reduction (CVR) on a National Level*, Pacific Northwest National Laboratory report, July 2010.

[13] A. C. Mahdi Kiaee, David Infield, 'Demand Side Management Using Alkaline Electrolysers within the UKGDS simulation network ', in The 21st International Conference and Exhibition on Electricity Distribution, Frankfurt, Germany, 2011.

Chapter 5 A hybrid model for residential loads with high PV penetration

5.1 Introduction

Electricity demand is stochastic process and varies with time. ADMD calculates the likely maximum demand for a group of customers within a specific period of time, the time duration of the received voltage is not considered [1], and it does not take into account the stochastic nature of the demand. ADMD should be statistically selected so there is an acceptable probability that actual load is less than the design figure. The LV network is usually designed to supply acceptable voltage at the time of greatest possible demand in the traditional approach, which assumes every consumer simultaneously uses every appliance. However, in real life consumers do not reach their maximum loads at the same time, and their electricity demand reflects diversity. Weather and temperature fluctuation adds to the diversity of demand, too. Furthermore, many appliances controlled by thermostats, such as air-conditioning and refrigerators, also exhibits loads as the results of heat loss or gain. The impact of new technologies, such as an uptake of electric vehicles, roof top solar generation or an increased use of demand management is increasingly studied using probability based tools including Monte Carlo based methods. Statistically representative feeder test cases and load models are required if the simulation results are to be reflective of the actual network performance outcomes. There are several factors usually being taken account into demand predicting and modeling, such as time period, weather and customer type.

This chapter focuses on a new hybrid approach for residential load modelling that seeks to separately capture the underlying low frequency aggregation behaviour of load and the higher frequency stochastic behaviours.

5.2 The Pavetta low voltage network

The Australian Government's Perth Solar City program are designed to identify and understand barriers to the uptake of energy efficiency and renewable energy in the residential sector, test new energy efficiency technologies and undertake trials, and inform future government policy. The areas engaged in the Perth Solar City program are shown in map in Figure 5.1.

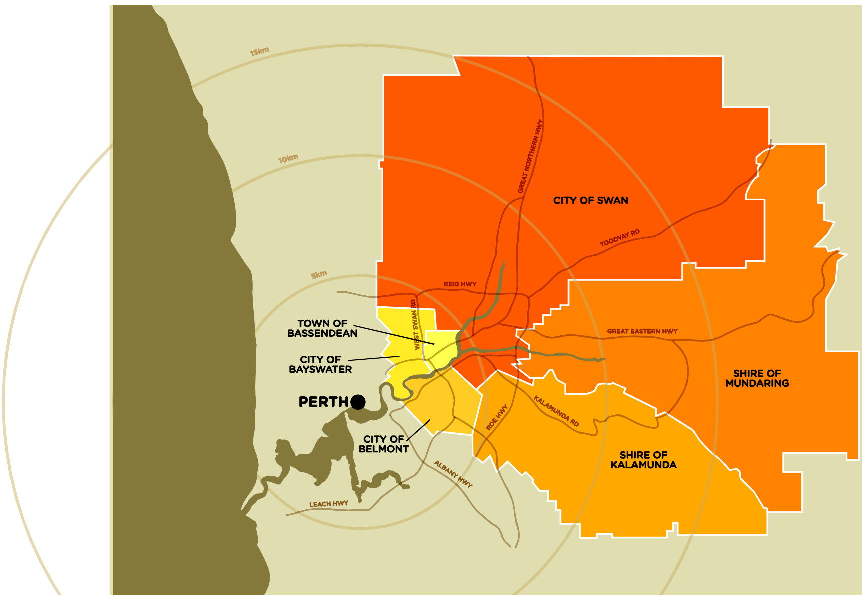


Figure 5.1 Map showing the areas engaged in the Australian Government's Perth Solar City program.

Western Power Corporation conducted a high photovoltaic (PV) penetration trial with funding from the Australian Federal Government Solar Cities Program. As part of the Perth Solar City program, the PV trial is to study impact of high penetration PV on the distribution network:

- Provide research data and information to validate simulation models looking at how a saturated distribution network performs;

- Investigate Power Quality issues that might occur as a result of increasing penetration of PV systems;
- Investigate what conditions on the network have an impact on the performance of customers' PV systems;
- Evaluate potential for distributed PV generation to provide network support;
- Provide recommendations to maintain network performance standards under existing and forecast PV uptake trends.

The pole mounted Pavetta transformer in the suburb of Forrestfield was chosen prior to this study. This is a 230V/400V three phase four-wire LV feeder supplied from a 200kVA Dyn transformer. The reason it needed to be in PSC catchment area was to ensure availability of customer level smart meter data. The smart meter rollout in the PSC catchment area allowed the wide ranging monitoring of network performance at the customer level.

TABLE 5.1
Summary of the PV Information Serviced by Pavetta Transformer in the Suburb of Forrestfield

Customers Connection Phase	No. of Customers	No. of PV Customers	PV Capacity Installed (kW)	PV Penetration by No. of Customers	PV Penetration by Transformer Capacity
3 Phase	26	14			
Red	14	4	20.2	16%	10.1%
White	17	4	15.0	10%	7.5%
Blue	20	12	23.4	18%	11.7%
Total	77	34	58.6	44%	29.3%

The trial provided an opportunity to collect consumer data in a high solar penetration environment for cluster analysis and load modelling. The summary of the residential customer information are listed in Table 5.1. Among the 77 customers as shown in Figure 5.2, there are 8 PV customers before trial, 20 PV customers recruited for the trial, and 6 PV customers after the trial. A total of 54kWp of roof top PV was installed by 34 customers with the typical ratings of 1.88 kW. The locations of the residential customer in the PV trial are shown in Figure 5.3. Table 5.1 shows the summary of the PV information serviced by the Pavetta transformer in the suburb of Forrestfield. Table 5.2 shows the summary of the average daily energy consumption of the 77 customers serviced by the Pavetta transformer in Forrestfield.

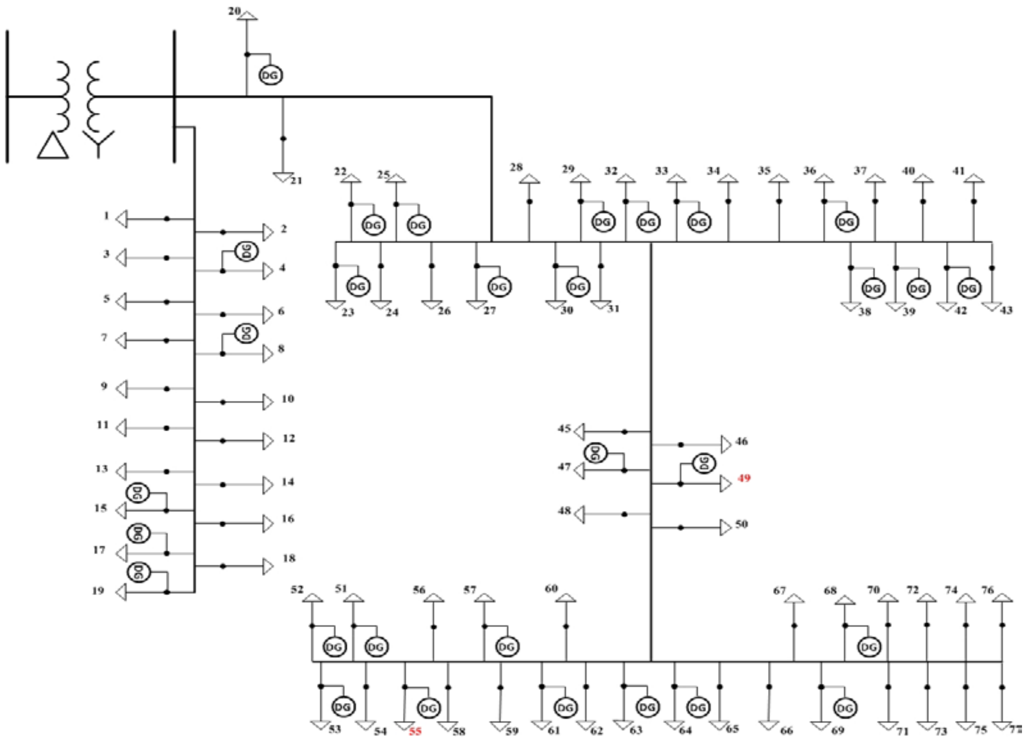


Figure 5.2 A scheme showing the 77 the residential customer engaged in the high penetration PV trial.

TABLE 5.2
Summary of the Average Daily Energy Consumption of the 77 Customers Serviced by Pavetta Transformer in the Suburb of Forrestfield (kWh)

Customer Type	**Spring 2011**	**Summer 2011/12**	**Autumn 2012**	**Winter 2012**	**Yearly**
Non-PV	13.0	16.69	14.5	16.75	15.24
PV	12.21	17.73	14.03	18.42	15.60
Difference	-6.1%	6.2%	-3.2%	-10.0%	2.4%

Figure 5.3 Map showing the locations of the households engaged in the PV trial

Smart meters have been installed, which enables to see the load and voltage profiles at the customer level. Consumer smart meter data have been made available on 15 minute intervals on import and export energy, voltage and currents began in July 2011. Distribution transformer is fitted with a logger in 5 minute intervals. Recorded load data of 2 of the customers were invalid, so there are in total 75 customers under the load study.

Chapter 4 reviewed a database of physical LV feeders (230V/400V three phase) in the Western Power network using statistical cluster and discriminant analysis [2]. It was shown that the Perth LV distribution feeders could be represented by eight prototypical feeders as shown in Table 4.12 in Chapter 4.

The Perth Solar City feeder, Pavetta Street, belongs to Cluster Two - Residential with mainly using overhead lines. The feeder was selected prior to the conduct of the cluster analysis as being representative of older overhead conductor systems that are expected to have higher sensitivity to PV induced voltage rise and unbalance. Figure 5.4 illustrates the Pavetta Street Feeder is close to the cluster centre in terms of the key physical parameter set and is a good representation of feeders of this type. There are 2,644 LV feeders in this cluster and the results will be transferrable to many other Perth feeders.

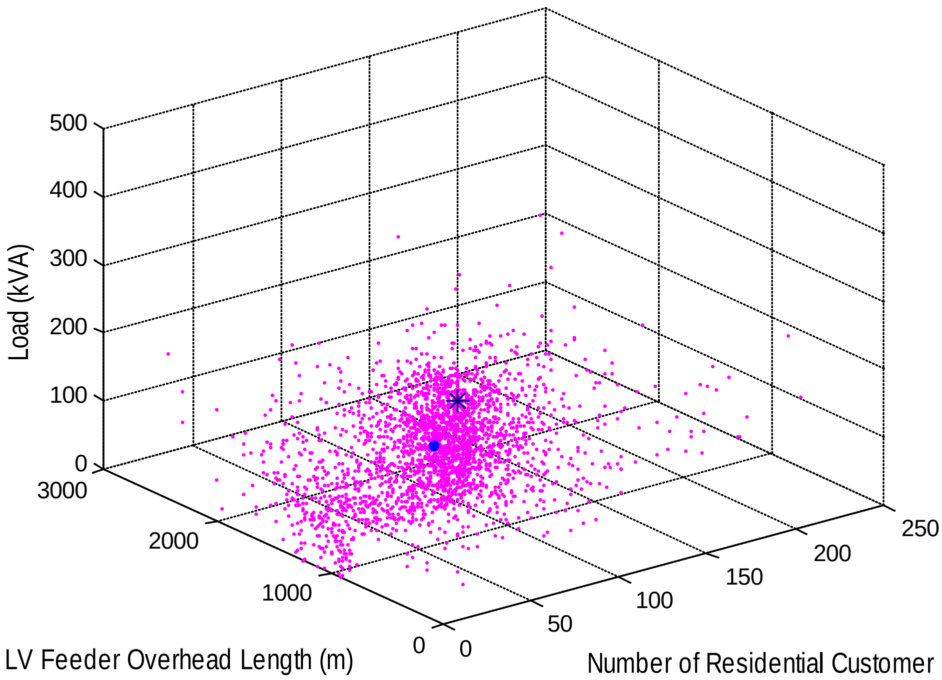


Figure 5.4 Plot illustrating the placement of the Pavetta Street feeder relative to the 2,644 LV distribution feeders in Cluster 2 in terms of the physical parameter set.

(The Pavetta Street feeder is marked by an asterisk, and the centre of cluster 8 is marked by a dot.)

5.3 Slow diurnal load variation and the statistical analysis for the high demand day model

The Pavetta Street households have smart meters that return 15 minute accumulated load energy measurements. This analysis focuses on high demand day modelling. Perth has a dominant summer air-conditioning peak and consumption is strongly correlated with temperature. In comparison, the modelling on low demand days will be implemented in the next chapter.

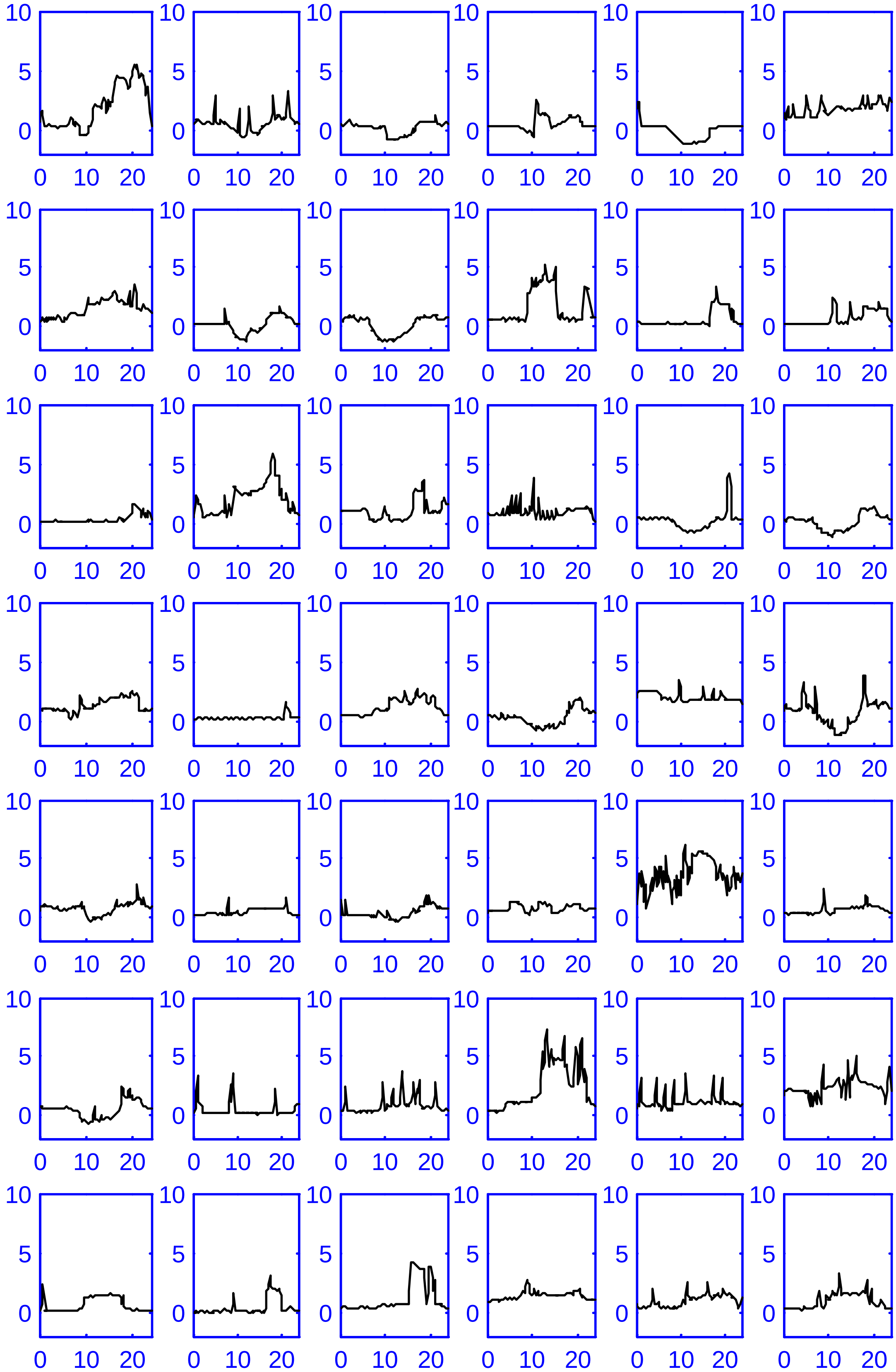
A number of stochastic process models for diurnal consumer demand have been published [1, 3-6]. This chapter seeks to identify specific load types within a set of residential load data to provide an improved understanding of the load structure. This is lost in the stochastic approaches. It will be shown that the lower frequency components of the load can be modelled using the Fourier series with a period of 24 hours. A cluster analysis can be applied to the Fourier coefficients to identify loads with similar daily consumption patterns. In this trial, six prototypical daily demand curves are identified that relate to certain load categories and behavioural patterns.

5.4 Classification with the refined analysis method

5.4.1 Data acquisition

Figure 5.5 shows a typical high-demand day load profile for each of the 75 customers. The individual residential electricity demand profiles have high variability in the daily energy consumption. Significant air-conditioning loads are visible. Many consumers display negative real power demands at midday due to PV generation. The aggregated demand curve, shown in the bottom right is much smoother due to diversity in the operation of consumer loads.

The high penetration photovoltaic trial is subjected to local weather conditions, such as sunshine hours and daily global solar exposure. Residential consumers use air-conditioners more when the daily maximum temperatures are high. Hence, demand data on the ten hottest days, picked according to its daily maximum temperature, were used in load modelling. The meteorological database was provided by Australian Bureau of Meteorology. Station Number 9021 (Airport) was used, as this station is close to Pavetta Street.



[cont.]

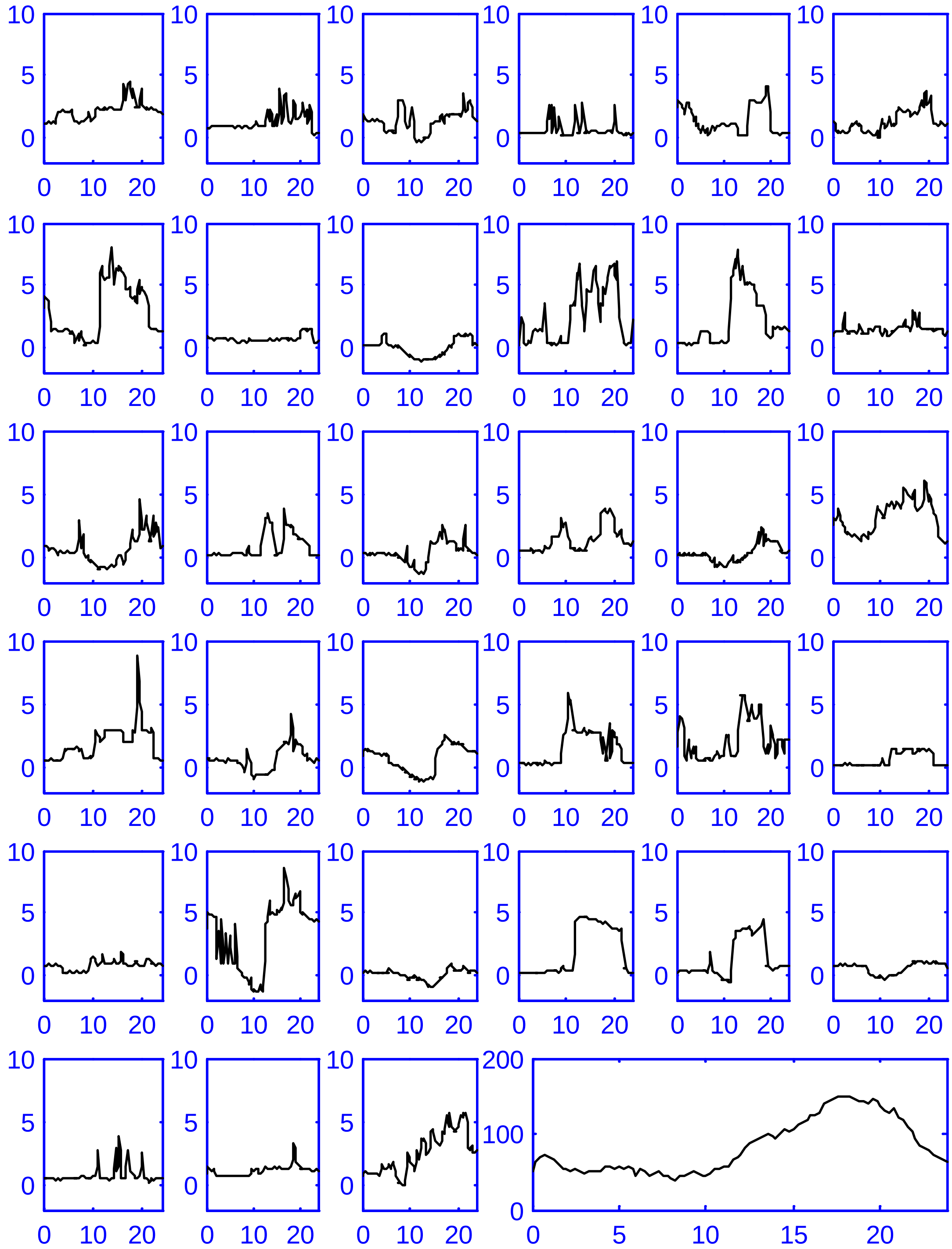


Figure 5.5 A typical daily load profile of each 75 customers on Pavetta Street, with the last subfigure showing the aggregated load profile of 75 customers. (x label: Time (Hour); y label: Load (kW))

During the photovoltaic trial from June 2011 to June 2012, the top two daily maximum temperatures were 41.5°C (26th January 2012) and 41.0°C (28th January 2012), however, the data on the two days were not included in the ten hot summer days, because the database missed several households' data on those two days. The feeder data of customer No. 510004921 were not recorded after 7:30am on 26/01/2012, and the feeder data of tens of customers were not recorded after 6:30pm on 28/01/2012.

Also, the data on a specific day will not be analysed if the data of the day missed more than three continuous time points (including three) for any of the seventy-five households under study. The missing data on valid days were linear interpolated before Fourier Transform.

The maximum temperature of the ten days under study in the load modeling varied from 38°C to 40.9°C as listed in Table 5.3. Table 5.3 also listed the meteorological information relevant to electricity consumption/generation on the ten days on rainfall amount, minimum temperature, daily global solar exposure, evaporation, sun-hours, temperature at 9am, relative humidity at 9am, temperature at 3pm and relative humidity at 3pm. Load data on the five days (i.e. 28th December 2011, 25th January 2012, 5th March 2012, 9th March 2012 and 11th March 2012) were selected to build the load model, while data on the other five days (i.e. 24th January 2012, 10th February 2012, 6th March 2012, 10th March 2012 and 12th March 2012) were put aside to be used as test cases for model verification.

This load modeling helps to support investigation into the effects on LV urban electricity networks of future wide-scale uptake of solar technologies. The modeling may contribute to determine typical customer loading profile, improve feeder analysis, design the tariff program and provide more detailed information for consumers (e.g. peak demand time and environmental information).

TABLE 5.3
Summary of Meteorological Information on the Ten Days with High Maximum Temperature

Date	Maximum temperature (°C)	Minimum temperature (°C)	Daily global solar exposure (MJ/m^2)	Evaporation (mm)	Sun-hours (H)	Rainfall amount (mm)	Temperature at 9am	Relative humidity at 9am (%)	Temperature at 3pm	Relative humidity at 3pm (%)	Daily maximum aggregated load (A)
2011/12/28	38.8	19.7	29.6	12.0	10.9	00.0	28.4	29	38.1	7	686.0
2012/01/24	38.0	20.7	29.4	10.0	11.9	00.0	25.1	53	37.0	26	826.4
2012/01/25	40.0	24.6	29.9	15.0	13.3	00.0	31.5	39	39.4	17	818.8
2012/02/10	40.3	23.2	26.3	09.2	12.6	00.0	28.7	37	39.7	14	684.9
2012/03/05	39.2	15.2	24.9	09.2	11.7	00.0	29.8	20	36.4	11	643.3
2012/03/06	38.8	18.3	24.7	11.6	10.6	00.0	31.6	12	37.8	7	685
2012/03/09	38.3	17.6	24.2	08.8	11.6	00.0	23.2	47	37.2	16	687.2
2012/03/10	40.4	15.6	24.4	08.0	11.3	00.0	30.9	24	37.3	10	637.2
2012/03/11	40.9	18.5	23.7	11.6	11.3	00.0	27.1	45	40.5	8	710.0
2012/03/12	40.4	14.5	23.3	09.2	11.3	00.0	30.6	19	34.4	22	710.7

5.4.2 Discrete Fourier transform (DFT) based load analysis

DFT algorithm was used to determine the Fourier series coefficients for each consumers daily load [7]. The DFT of f_n is a sequence of N complex numbers given by:

$$F_k = \sum_{n=0}^{N-1} f_n e^{-i2\pi \frac{k}{N} n} \tag{5.1}$$

The Inverse DFT (IDFT) is given by:

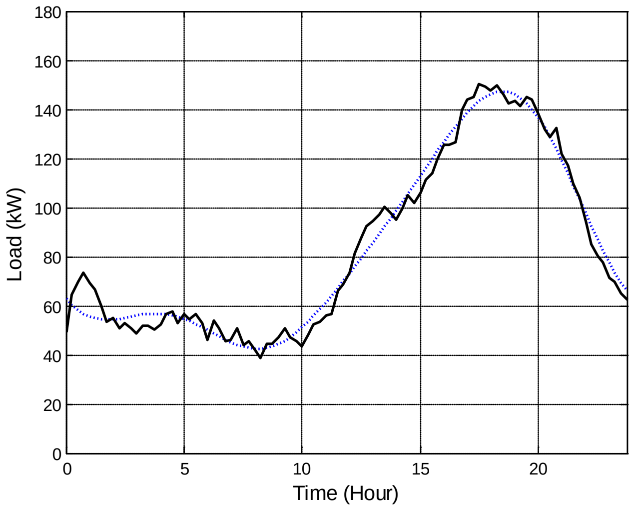
$$f_n = \frac{1}{N} \sum_{k=0}^{N-1} F_k e^{i2\pi \frac{k}{N} n} \tag{5.2}$$

The truncated Fourier series is:

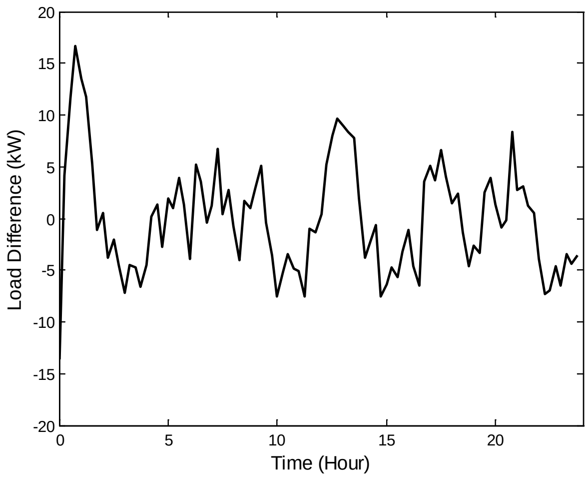
$$f_{n\_LF} = \frac{A_0}{2} + \sum_{k=1}^{N-1} \left[A_k\, cos(\frac{2\pi kn}{N}) + B_k\, sin(\frac{2\pi kn}{N}) \right] \tag{5.3}$$

The Fourier representation was truncated to complex coefficient vectors of length of four, from A_0 to A_3+iB_3. The Inverse DFT was used to produce a low frequency Fourier series model for each consumer load profile. This approach is demonstrated in Figure 5.6. The solid line in Figure 5.6 (a) shows the daily aggregated load profile on the Pavetta Street feeder while the dashed line shows the low frequency model. The daily load difference between the empirical and IDFT curves is illustrated in Figure 5.6 (b).

The selected five load data for building model were incorporated into one, as the accuracy level improves when more load data used to build the model. Then a finite sequence of real or complex numbers of the DFT were extracted for the purpose of cluster analysis. The more sequence of real or complex numbers extracted of DFT, the smaller load difference between the empirical daily aggregated load and DFT curves is, yet the more cluster parameter needed to be dealt with and less efficiency in clustering.



(a) Solid line showing a typical daily aggregated load profile of 75 customers on Pavetta Street. Dotted line showing the DFT/IDFT fitted load curve with the vectors of length of four.



(b) The load difference between the empirical (solid line in (a)) and IDFT (dotted line in (a)) curves.

Figure 5.6 Graphs demonstrating the approach in which Inverse DFT was used to produce a low frequency Fourier series model for each consumer load profile.

Figure 5.7 illustrates the variation in the sum of squared error, between the physical and the low pass filtered model for increasing Fourier coefficient vector length, which illustrate the way in choosing the optimal number of sequence of real or complex numbers. Longer Fourier vectors reduce the differences between the empirical daily aggregated load and DFT curves but causes difficulty in the clustering process which aims to group consumers with similar diurnal behavior. Figure 5.7 shows a rapid reduction in total squared error for vector lengths up to four. Above that the gains are incremental and made at the expense of reduced effectiveness in the clustering stage. A vector length of four is the optimal choice.

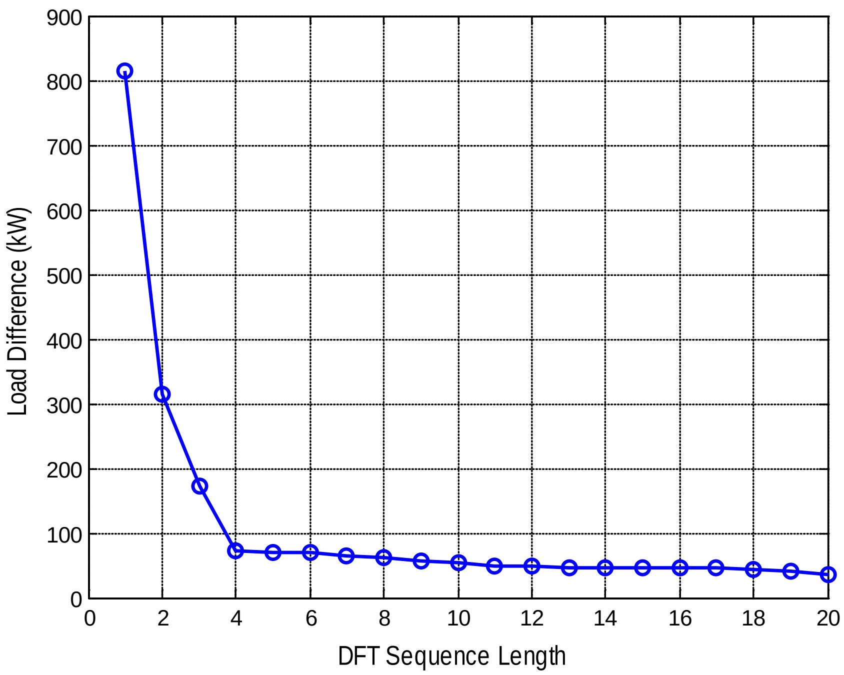
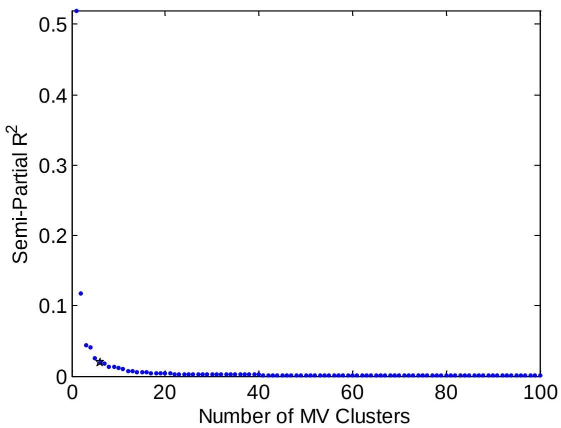


Figure 5.7 Curve illustrating the load difference decreases with the increasing DFT sequence length.

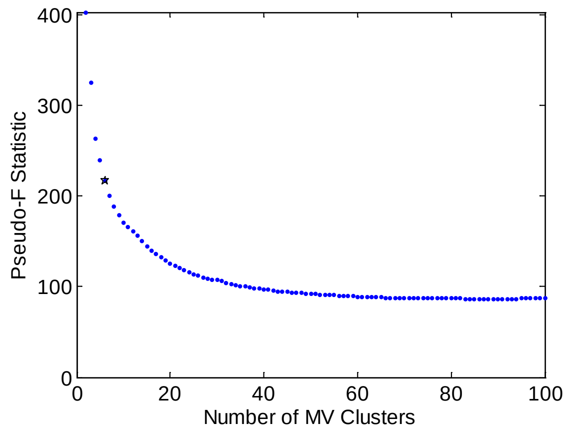
After using the DFT to extract values for A_0, A_1, A_2, A_3, B_1, B_2 and B_3, clustering process was carried out using a combination of hierarchical and discriminant cluster analysis. To produce a large enough data set to conduct a clustering analysis, the 75 consumer profiles from each of the five days were combined to produce 375 consumer profiles.

5.4.3 Classification using hierarchical cluster analysis

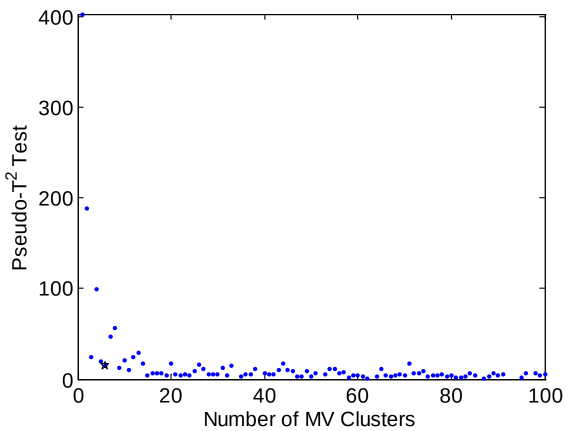
One crucial step in the cluster analysis is to select the optimal number of clusters with analysing the three statistical parameters of Semi-Partial R^2, pseudo-F statistic and pseudo-T^2 Test [8, 9], and the step has been introduced in Chapter 3. As shown in Figure 5.8, the favourable number of clusters could be no less than five according to *SPRSQ*. The pseudo-F suggests no more than twenty, and the pseudo-T^2 recommends three, six, or nine. The optimal grouping number was found to be six (m_{opt} = 6), so the 375 load profiles under study were separated into six clusters, each representing a type of load profile and electricity usage habit.



(a)



(b)



(c)

Figure 5.8 Graphs shows the method to determine the optimal number of clusters in the high demand day model, which is 6 (marked with ☆).

(a) – (c) showing the relationship between number of clusters and Semi-Partial R^2, Pseudo-F Statistic and Pseudo-T^2 Test, respectively.

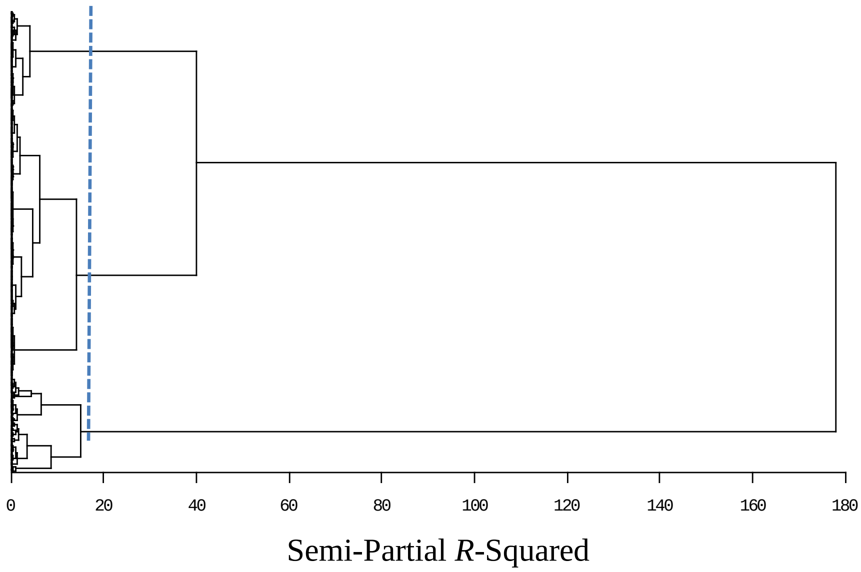


Figure 5.9 Dendrogram for load modelling for the high demand days resulting from MV feeders cluster analysis using the Ward's method.

Using the method introduced in Chapter 3, hierarchical clustering by Ward's method was chosen to categorise the load data as shown in Figure 5.9. Each load was firstly allocated to be a cluster on its own. The algorithm proceeds iteratively, at each stage joining the two most similar clusters, eventually until there is only one cluster remains. Ward's method uses the incremental sum of squares; that is, the increase in the total within-group sum of squares as a result of joining two different groups. At each stage, Euclidean distance between clusters is calculated to find out similarities between clusters [10].

5.4.4 Classification with the refined analysis method

Hierarchical cluster analysis was followed by a discriminant clustering analysis step to produce a stable cluster classifier.

A set of discriminant functions can be extracted and be used as the classifier for any test cases to determine to which group an individual case belongs as expressed in Eq. (3.14), where, $k = 1,\ldots,m_{opt}$; and $d = 7$. x_i and x_j (i, $j = 1,\ldots,7$) are the Fourier coefficients A_0, A_1, A_2, A_3, B_1, B_2 and B_3. The coefficients of a_0, a_i and a_{ij} for each cluster are listed in Table A.4 in Appendix. Six calculated values (i.e. $G_1,\ldots,G_6$), are determined for each profile using Eq. (3.14). For example, using Eq. (3.19), the QDF of Cluster 1 can be written as:

$$\begin{aligned}
G_{k=1} = {} & a_0 + a_1 A_0 + a_{11} A_0 A_0 + a_{12} A_0 A_1 + a_{13} A_0 A_2 + a_{14} A_0 A_3 + a_{15} A_0 B_1 + a_{16} A_0 B_2 + a_{17} A_0 B_3 \\
& + a_2 A_1 + a_{22} A_1 A_1 + a_{23} A_1 A_2 + a_{24} A_1 A_3 + a_{25} A_1 B_1 + a_{26} A_1 B_2 + a_{27} A_1 B_3 \\
& + a_3 A_2 + a_{33} A_2 A_2 + a_{34} A_2 A_3 + a_{35} A_2 B_1 + a_{36} A_2 B_2 + a_{37} A_2 B_3 \\
& + a_4 A_3 + a_{44} A_3 A_3 + a_{45} A_3 B_1 + a_{46} A_3 B_2 + a_{47} A_3 B_3 \\
& + a_5 B_1 + a_{55} B_1 B_1 + a_{56} B_1 B_2 + a_{57} B_1 B_3 \\
& + a_6 B_2 + a_{66} B_2 B_2 + a_{67} B_2 B_3 \\
& + a_7 B_3 + a_{77} B_3 B_3
\end{aligned} \tag{5.4}$$

where, A_0, A_1, A_2, A_3, B_1, B_2 and B_3 are the Fourier coefficients. The 375 profiles are classified into six clusters according to the highest classification scores G_k.

After the initial discriminant analysis (DA) process, it has been found out that the classification results were not the same as the primary hierarchical cluster analysis, indicating the first DA had some unstable features. Clustering is a hierarchical single-pass process. Optimal classification requires, in general, back-tracking and multiple-passes. After seven iterative rounds of QDF re-substitutions, the classification error rate is zero, showing stable cataloguing has been achieved [11]. The results classified the 375 load profiles into 86, 155, 53, 40, 36 and 5 in each cluster, respectively.

5.5 Load modelling procedures

The Monte Carlo approach estimates unknown characteristics of the system via repeated experiments. This part of work adds a probabilistic dimension to the load modeling process to allow the production of load models that encapsulate more accurately the variability and underlying diurnal behavior of hot-day consumer loads.

5.5.1 Load model on the periodic components

For each of the six clusters, a histogram is constructed for the seven Fourier coefficients A_0, A_1, A_2, A_3, B_1, B_2 and B_3 and this is shown in Figure 5.10. To each histogram a normal distribution was fitted, as shown in Figure 5.11, and values calculated for the mean and standard deviation. Each line of subplots represents each of the six clusters. The fitted values of mean and standard deviation are shown as Table 5.4 and Table 5.5.

When constructing a load profile for a Monte Carlo simulation run the Fourier coefficients are selected by generating values using the normal distribution parameters. Randomly generated normal distribution points based on extracted values of standard deviation and mean are demonstrated in Figure 5.12. Figure 5.13 shows the low frequency periodic components for each of the six prototypical loads identified in the Pavetta Street feeder. Figure 5.14 shows the modeled high frequency components of load for each of the 75 customers on Pavetta Street.

TABLE 5.4
Values of Mean (μ) in Normal Distribution

	A_0	A_1	A_2	A_3	B_1	B_2	B_3
Cluster 1	1.1830	0.0543	-0.1287	-0.0157	0.3296	-0.0180	-0.0218
Cluster 2	0.5635	0.0657	-0.0979	-0.0163	0.1799	0.0463	-0.0164
Cluster 3	0.0964	0.3935	-0.1596	-0.0096	0.0536	0.1203	-0.0050
Cluster 4	1.9333	0.1557	-0.2069	-0.0259	0.5408	0.0970	-0.0027
Cluster 5	2.3456	-0.3878	-0.1167	-0.1130	0.8833	-0.1564	0.0324
Cluster 6	3.6853	-0.6108	0.0739	-0.1370	0.8873	0.1293	0.0690

TABLE 5.5
Values of Standard Deviation (σ) in Normal Distribution

	A_0	A_1	A_2	A_3	B_1	B_2	B_3
Cluster 1	0.2364	0.2652	0.2007	0.1208	0.2340	0.1373	0.1566
Cluster 2	0.2184	0.1870	0.1202	0.0851	0.1474	0.0974	0.0767
Cluster 3	0.1591	0.1215	0.0749	0.0669	0.0918	0.1083	0.0686
Cluster 4	0.3716	0.3326	0.3039	0.1234	0.3888	0.2003	0.2184
Cluster 5	0.3615	0.2669	0.2272	0.2207	0.2069	0.2322	0.2095
Cluster 6	0.3196	0.1731	0.2540	0.1748	0.3854	0.2477	0.1279

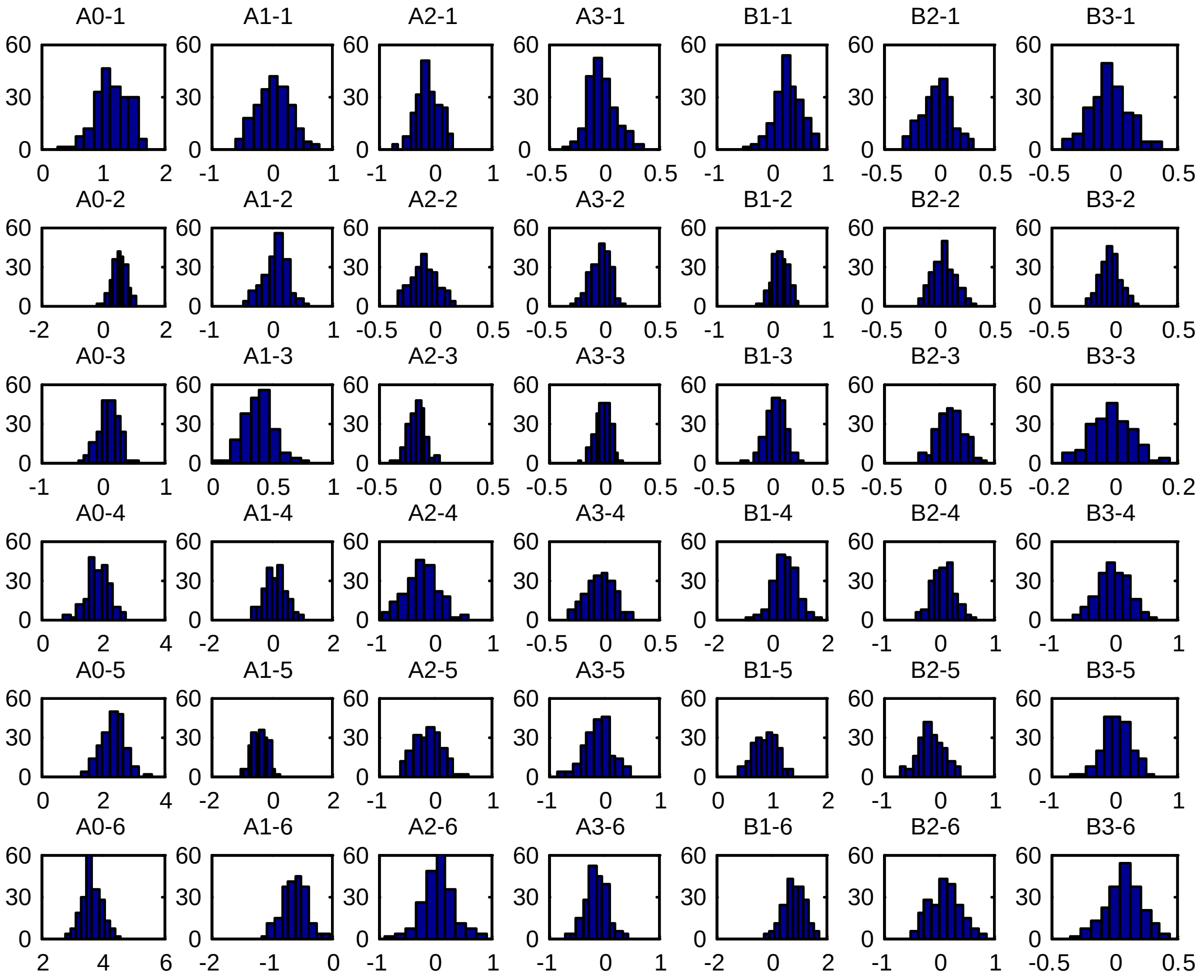


Figure 5.10 Histograms for A_0, A_1, A_2, A_3, B_1, B_2 and B_3 for each of the six clusters.

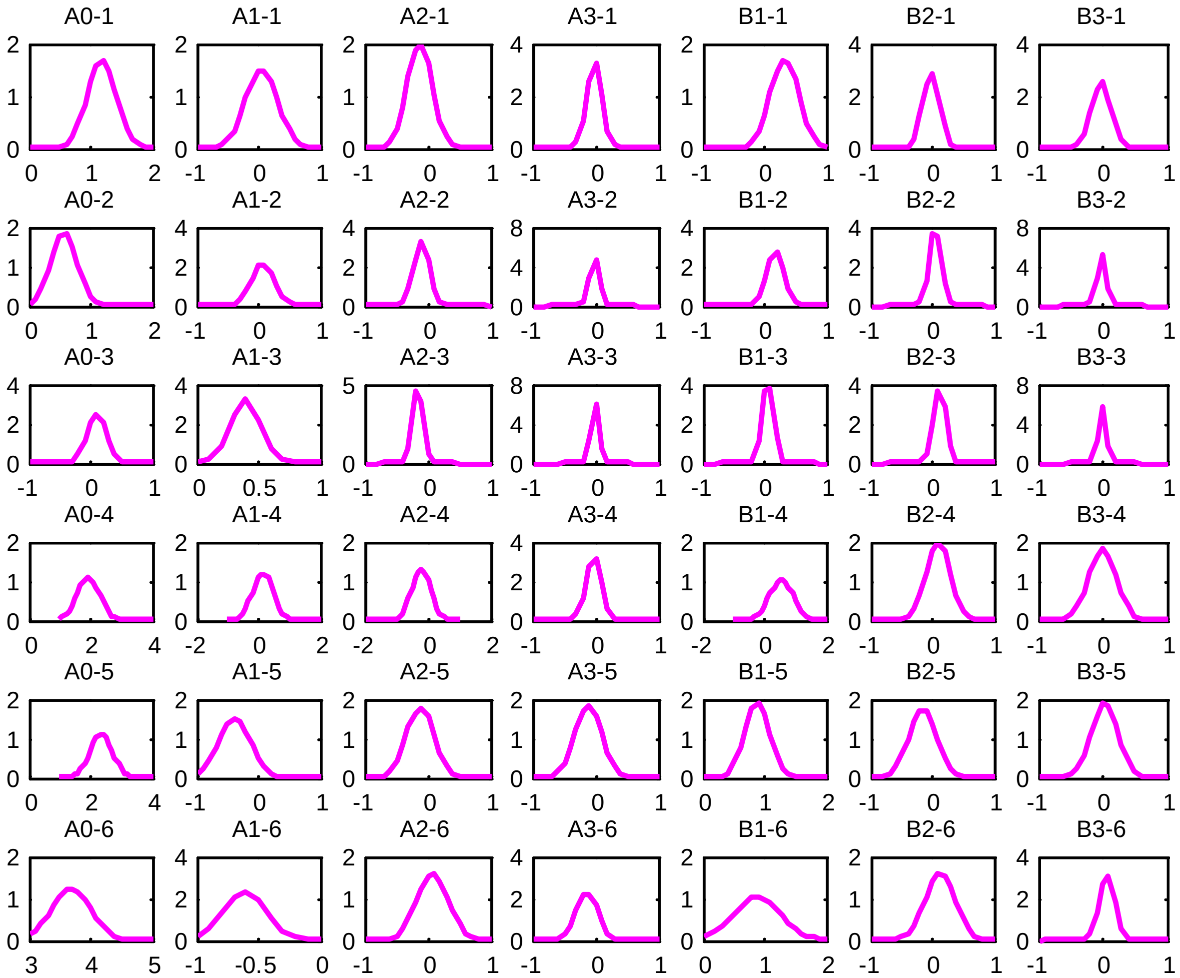


Figure 5.11 Computationally fitted normal distributions based on the histogram distribution of the Fourier components.

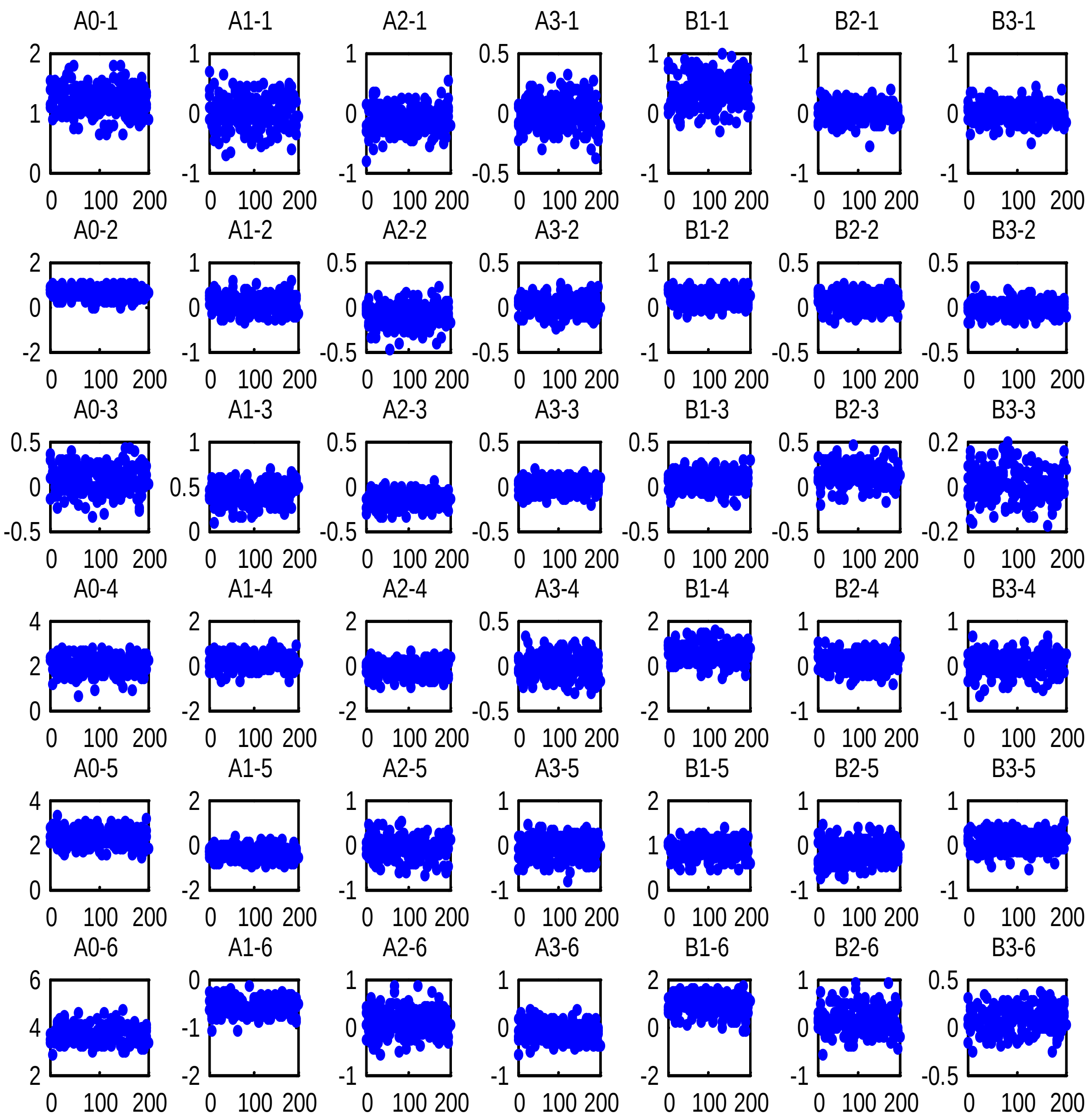


Figure 5.12 Randomly generated normal distribution points based on extracted values of σ and μ, with each line represents each cluster.

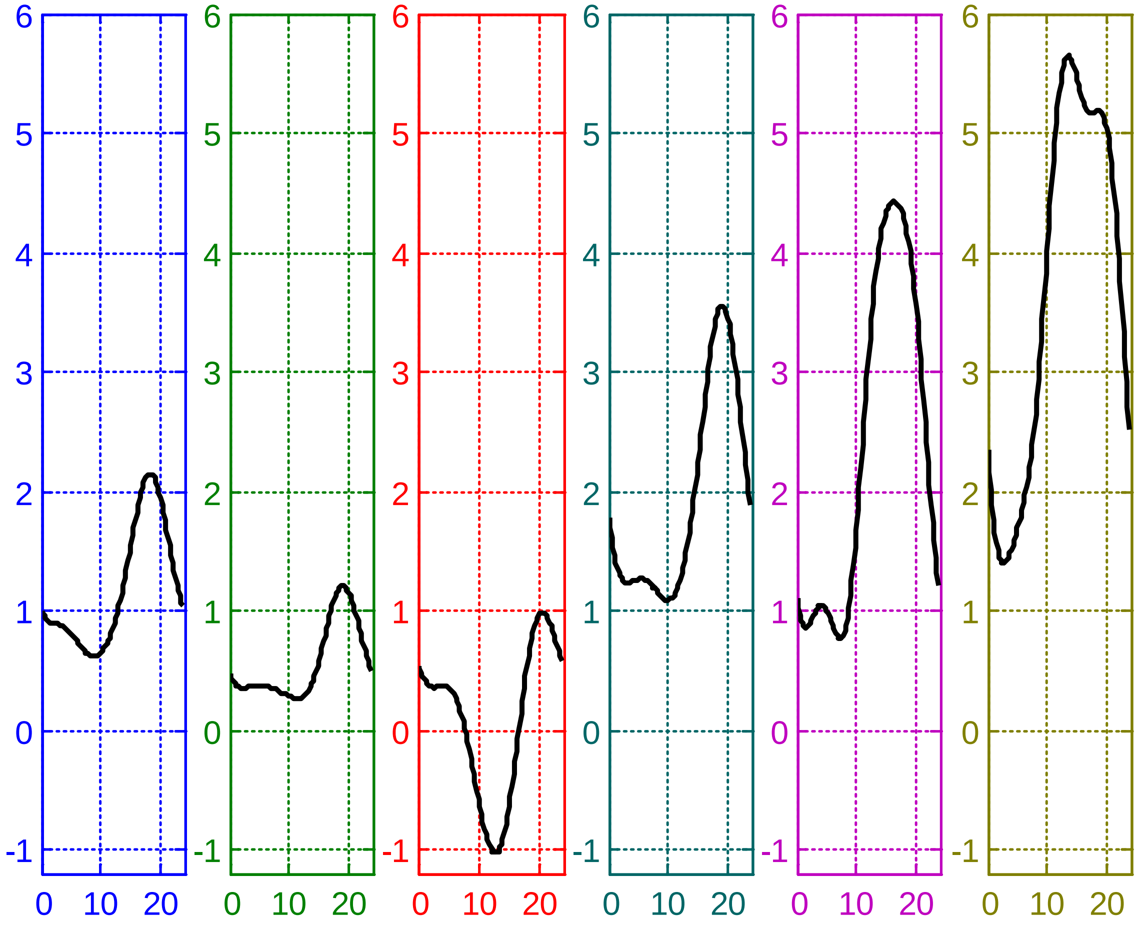
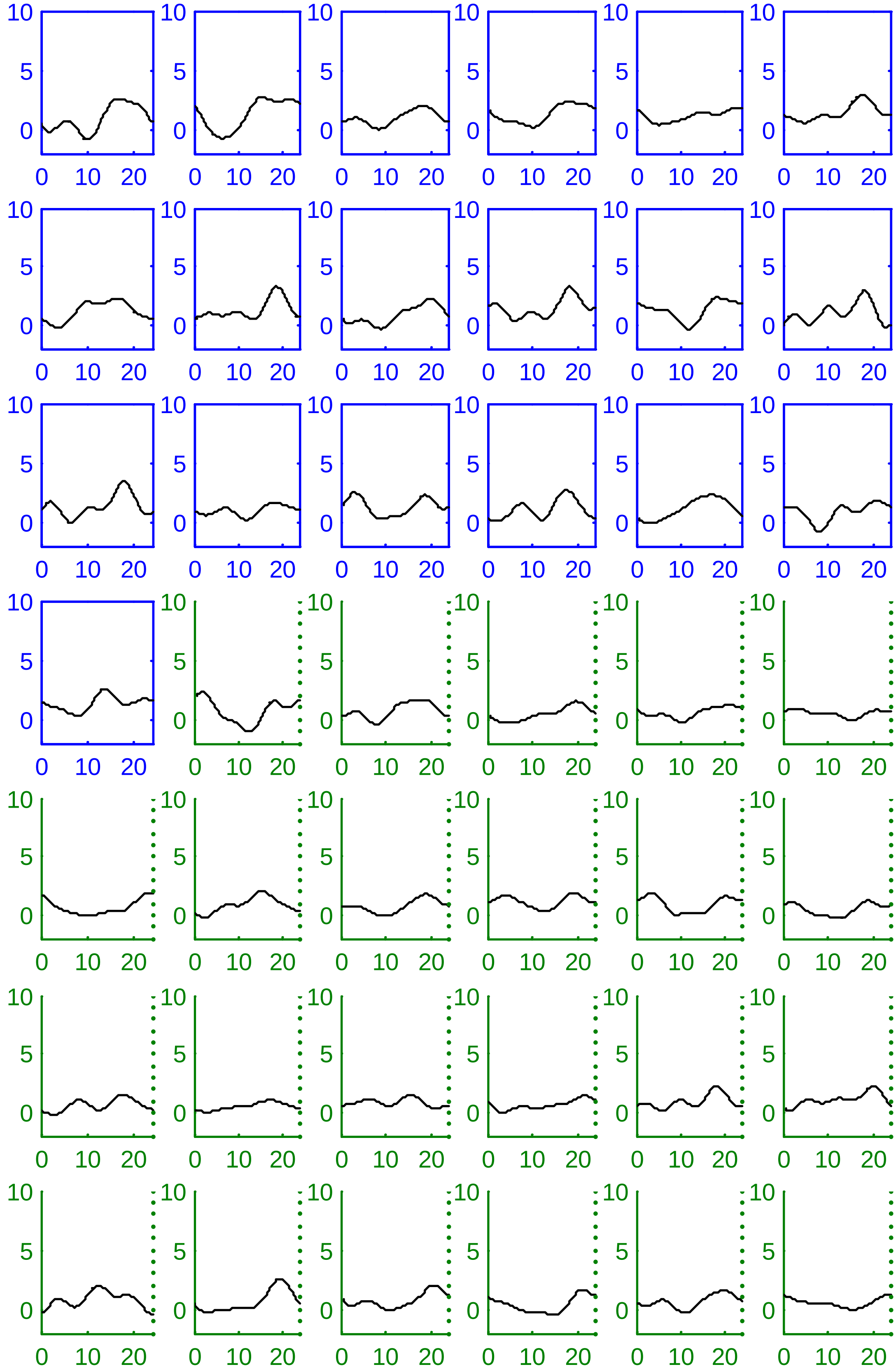


Figure 5.13 The periodic components of the six prototypical loads on Pavetta Street on high demand days (x label: Time (Hour); y label: Load (kW)).

Cluster one, 86 of the 375 profiles (23%), is a customer with moderate demand. Cluster two, containing 155 of the 375 profiles (41%), is the largest cluster and contains consumers with lower energy demands. The third plot, cluster 3, shows the highest penetration of domestic roof top solar systems. This represented 53 of 375 (14%) of the load cases studied. Clusters four and five, with 40 (11%) and 36 (10%) members respectively, are higher energy consumers. The sixth plot, cluster 6, shows consumers with the highest energy consumption patterns but only 5 of 375 (1.3%) profiles fell into this category.



[cont.]

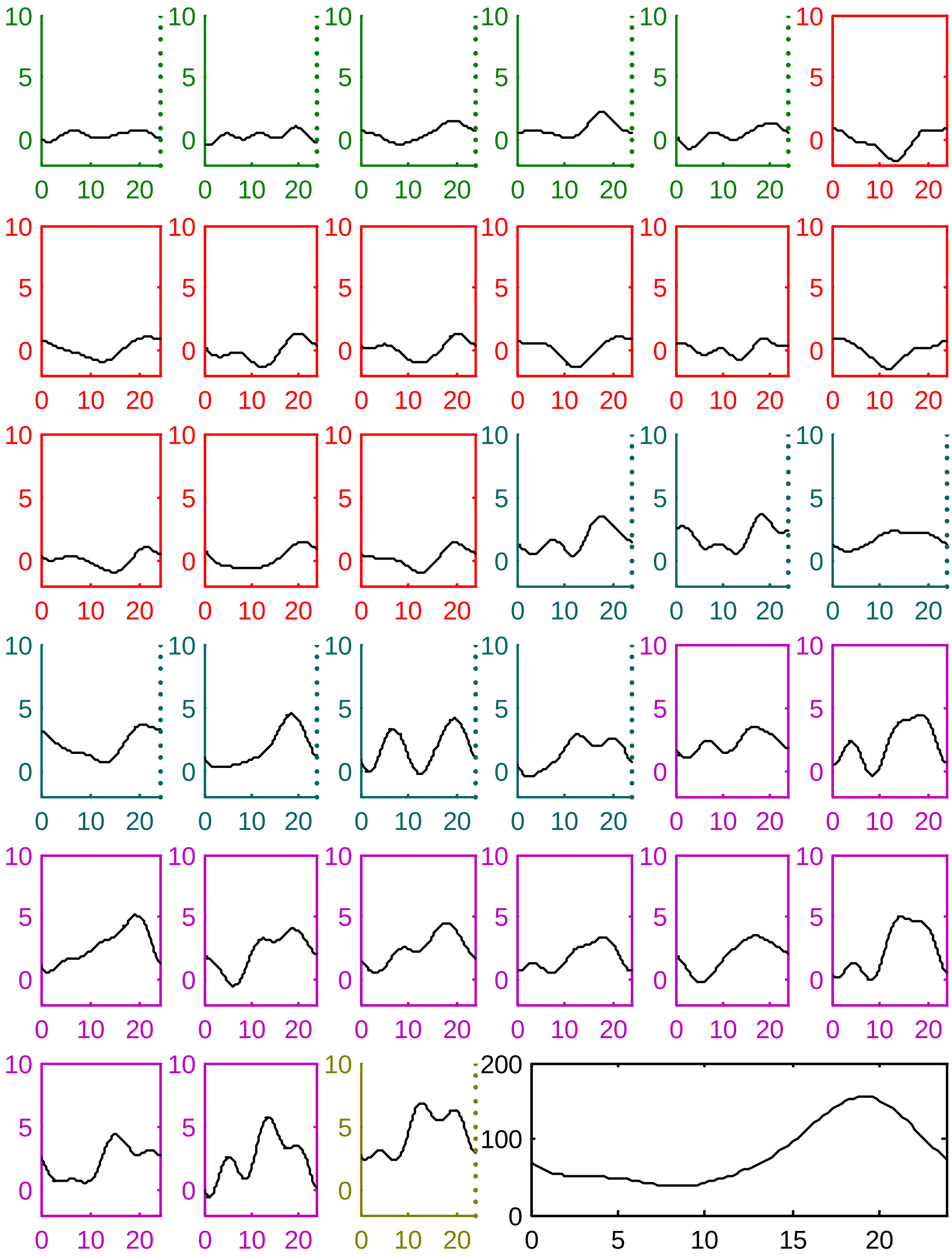


Figure 5.14 The modeled low frequency components of load for each of the 75 customers on Pavetta Street in the high demand day model.

The subplot boxes of cluster 1, 3, 5 use solid lines, and 2, 4, 6 use dashed lines. The aggregated curve is shown at the end subplot.

5.5.2 Load model for the high frequency components

The difference between the physical and the low frequency Fourier model for consumer load is, by definition a higher frequency component. The difference is given by:

$$\Delta f_{n_j} = f_{n_j} - f_{n_j\_LF} \tag{5.5}$$

Spectral density function of $\phi_r(f)$ for each cluster (r=1,…,6) is given by the following equation:

$$\phi_r(f) = \frac{1}{c_j}\sum_{j=1}^{C_j} \Delta F_{k_j}^2 \tag{5.6}$$

where, c_j represents the number of load within a specific cluster, which is, as mentioned previously, 86, 155, 53, 40, 36 and 5 for the six clusters, respectively.

This residual, is the result of fast variations in consumer power and largely the result of the switching of consumer loads. The high frequency terms are largely cancelled in load aggregation as the diversity in load switching time introduces random phase. An averaged power spectrum can be calculated for each of the six load profile clusters and is seen as Figure 5.15. Figure 5.16 shows the modeled high frequency components of load for each of the 75 customers on Pavetta Street.

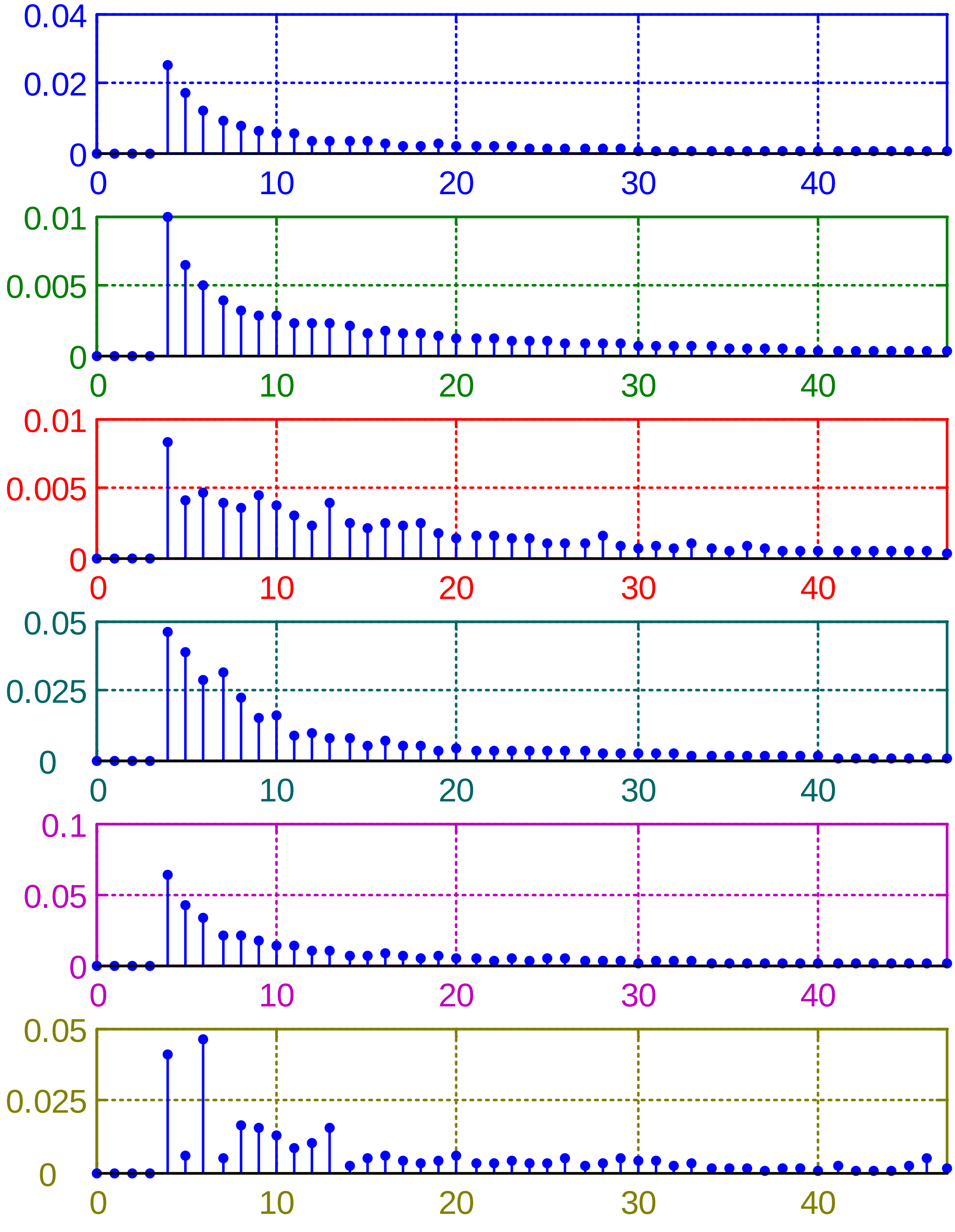
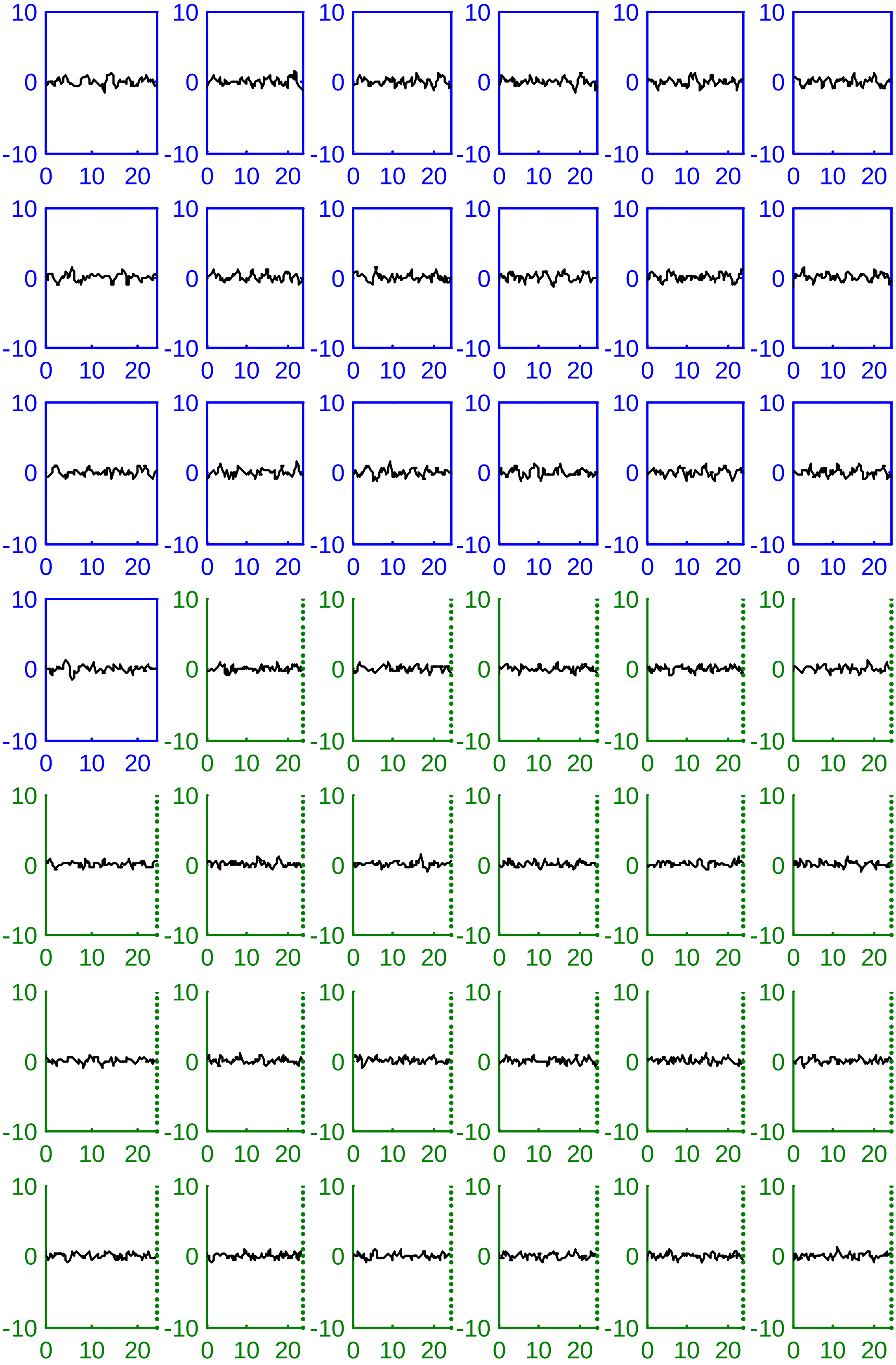


Figure 5.15 Frequency spectrogram for the six prototypical load profiles in the high demand day model. (x label: Frequency; y label: Averaged power spectrum)



[cont.]

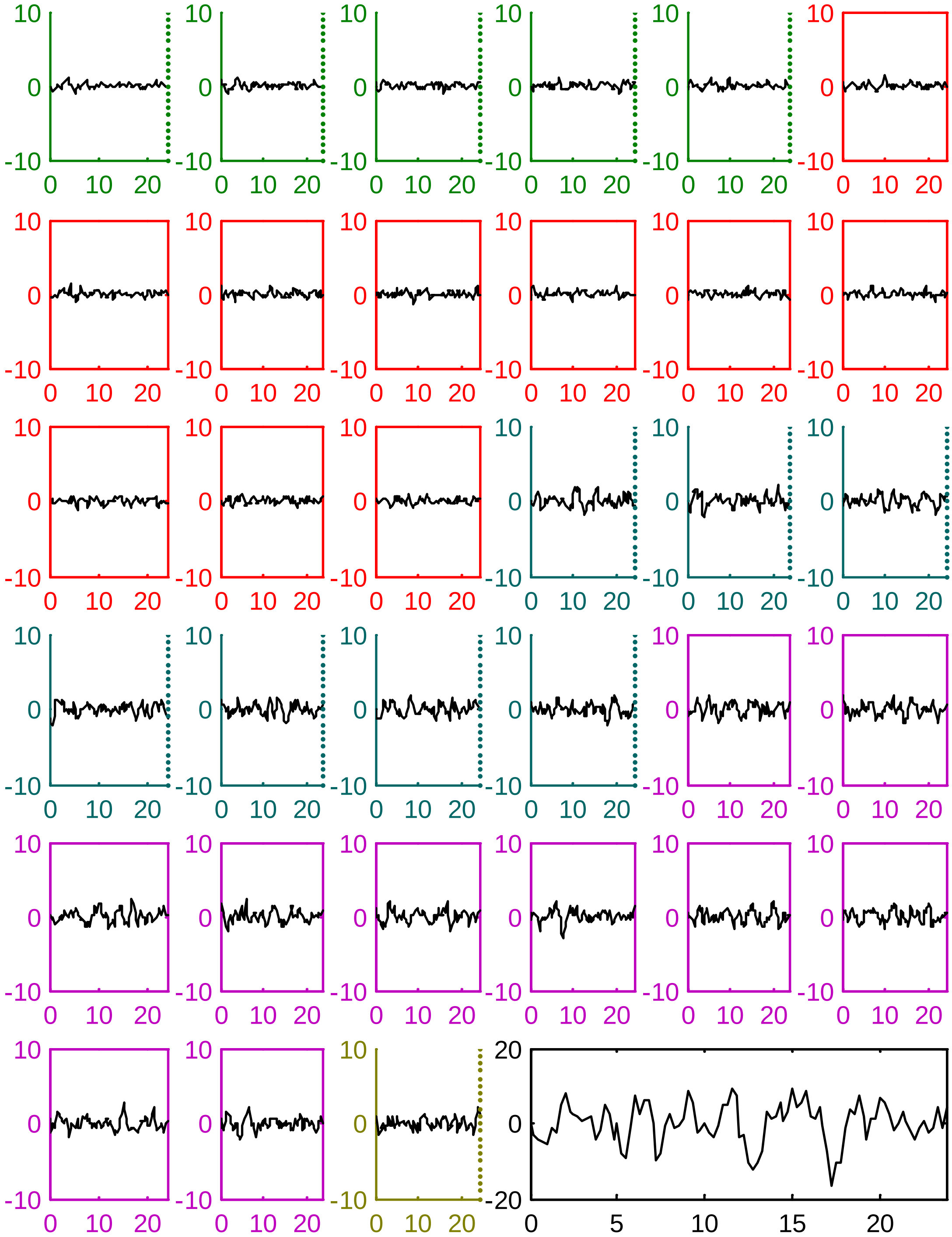


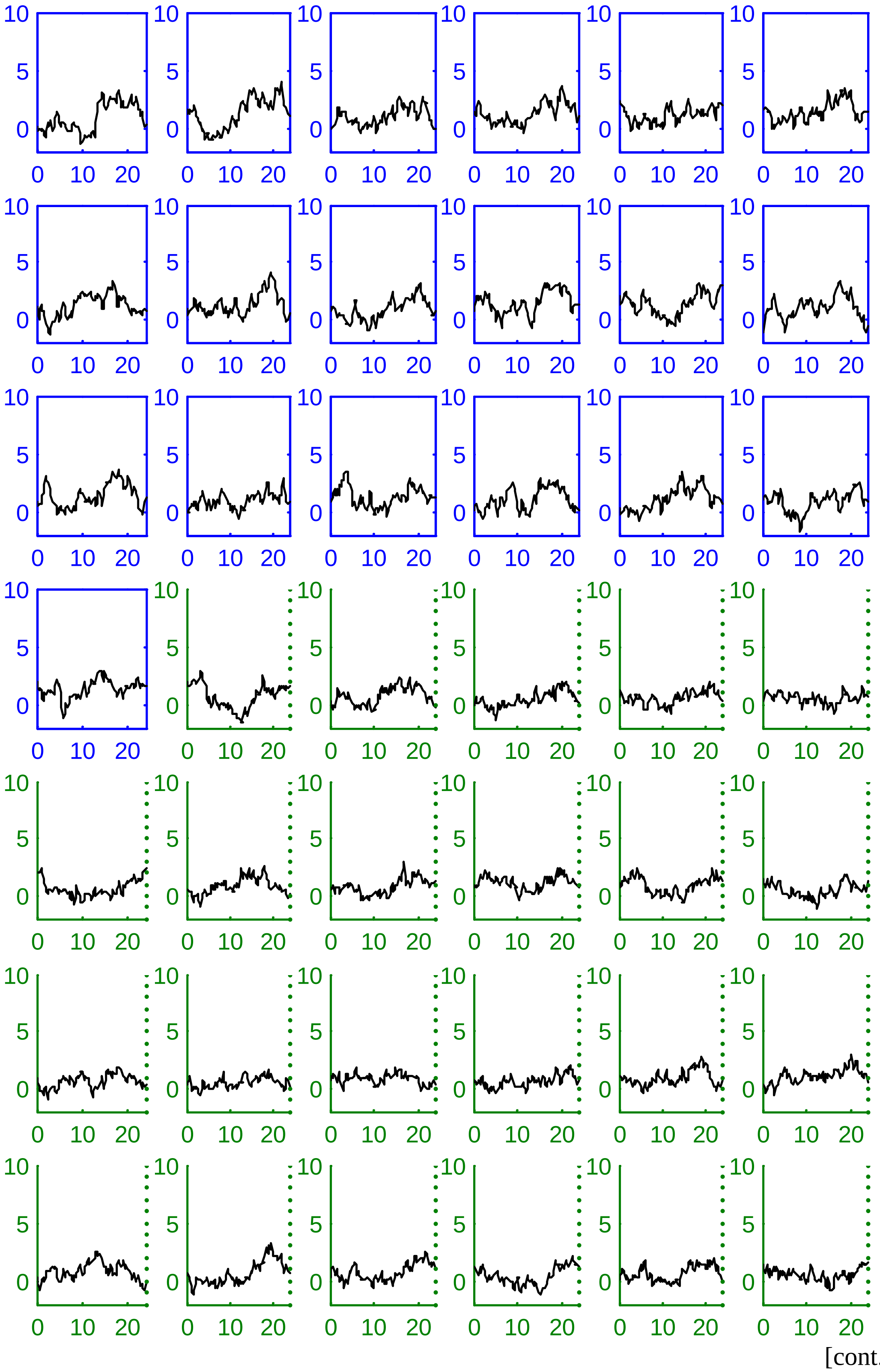
Figure 5.16 The modeled high frequency components of load for each of the 75 customers on Pavetta Street on high demand days.

The subplot boxes of cluster 1, 3, 5 use solid lines, and 2, 4, 6 use dashed lines. The aggregated curve is shown at the end subplot.

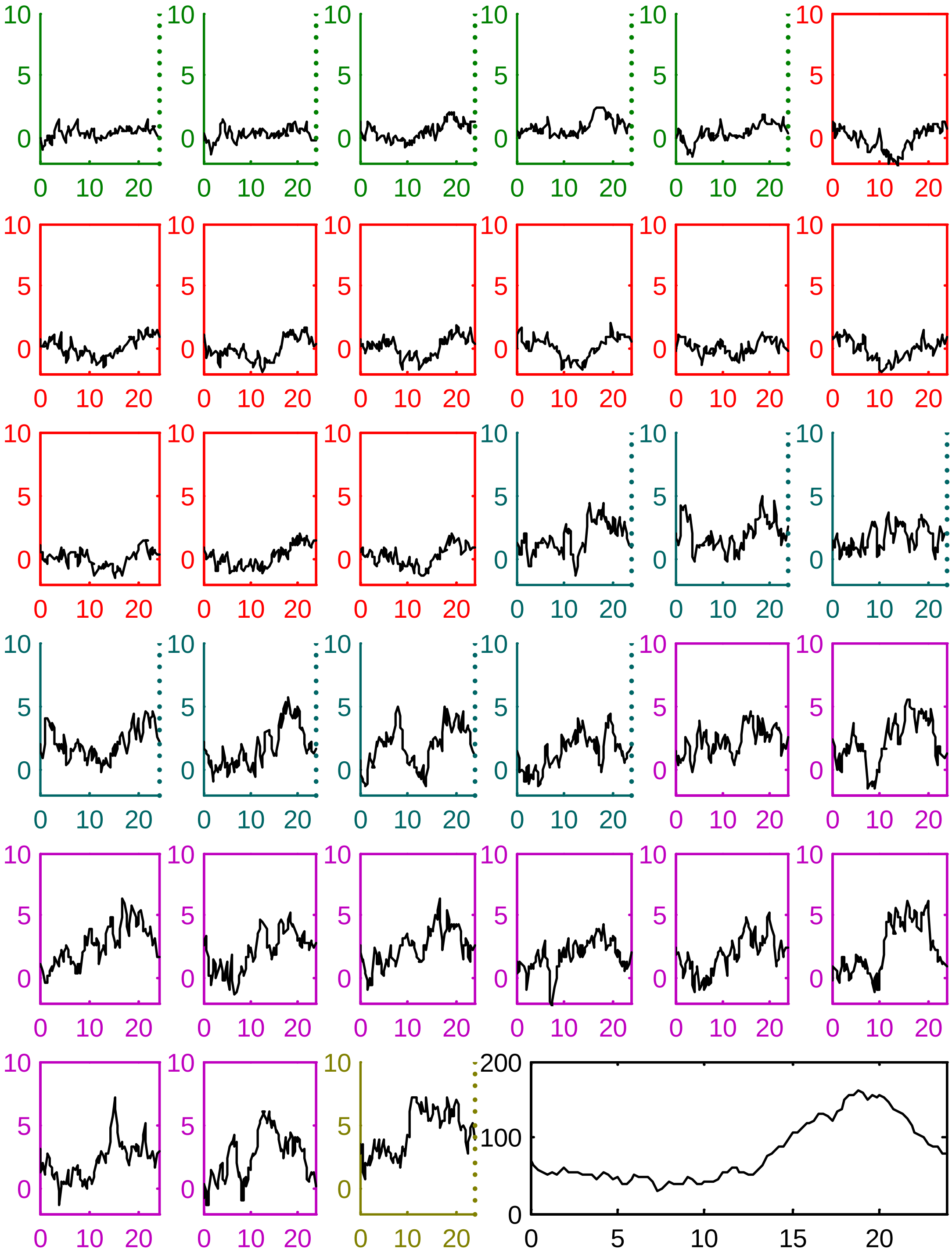
5.5.3 Load model results

For the purposes of constructing a model for a Monte Carlo simulation run, a randomised high frequency residual term is constructed by taking the magnitude of the average power spectrum terms and adding a random phase. A time sequence is readily generated using the inverse transform. An example of a final reconstructed load profile can be built by combining the low and high frequency components as seen in Figure 5.17. Its last subplot shows the modeled aggregated daily load. Figure 5.18 shows the reconstructed load profiles with the thicker curve showing the empirical load.

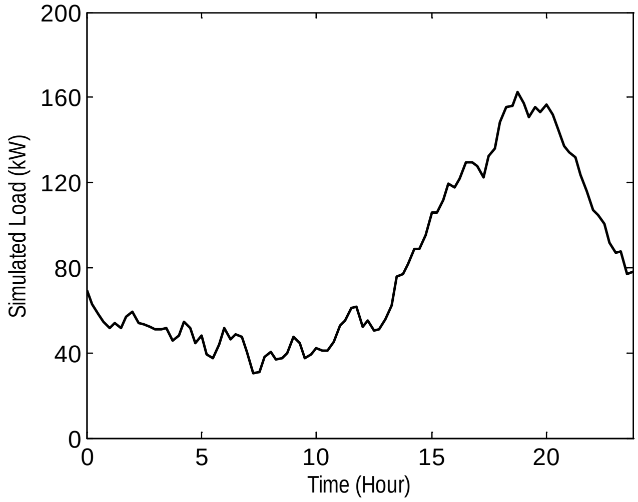
For the Monte Carlo based methods, repeated trials are required. A general consensus is that 100 trails are adequate, with more trials the higher accuracy. In choosing the trails, sampling interval and the repeated number, there are several issues that need to be considered, such as accuracy, parameterisation (i.e. network transferability) and simplicity. For a clearer view in the figure, only ten reconstructed load profiles are plotted in Figure 5.18. Figure 5.19 shows the ten times of reconstructed load profiles and the one hundred times in whisker plots, with the solid line showing the empirical load and the dashed line showing the mean of the reconstructed loads.



[cont.]



(a)



(b)

Figure 5.17 (a) The modeled low and high frequency components of load for each of the 75 customers on Pavetta Street in the high demand day model. The subplot boxes of cluster 1, 3, 5 use solid lines, and 2, 4, 6 use dashed lines. The aggregated curve is shown at the end subplot in (a), and an enlarged plot shown in (b).

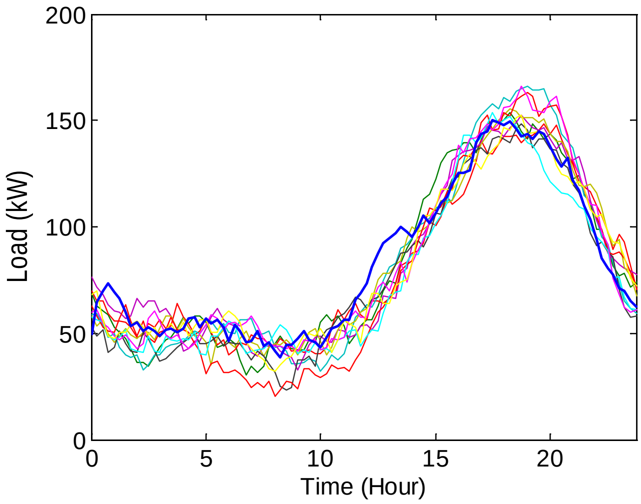
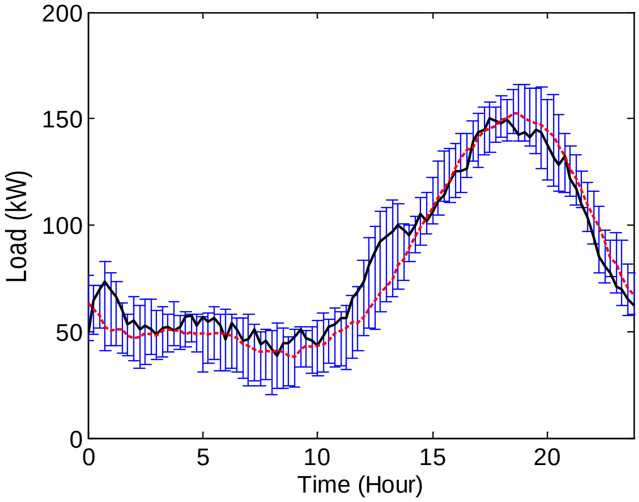
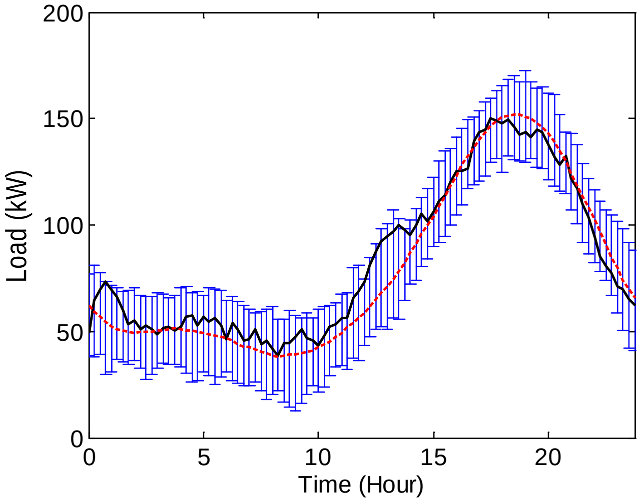


Figure 5.18 Ten reconstructed load profiles for the high demand days with the thicker curve showing the empirical load.



(a)



(b)

Figure 5.19 Whisker plots showing the reconstructed load profile for the high demand days with the solid line showing the empirical load and the dashed line showing the mean of the ten times of reconstructed load and one hundred times of reconstructed load, respectively.

5.6 Load model verification for the high demand day model

Five validation days (i.e. 24th January 2012, 10th February 2012, 6th March 2012, 10th March 2012 and 12th March 2012) were kept as test cases, while the other five days (marked with shading in Table 5.6) were combined to build the load models in the previous sections. The models presented are single residence models which have high variability. A key requirement is accurate modelling of load aggregation behaviour. This is assessed by assembling an ensemble of single residence loads, reflective of the test day, and comparing the group aggregation behaviour. This is listed in Table 5.6.

TABLE 5.6
Summary of Cluster Membership on the Ten Days

Date	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
28/12/2011	17	31	12	9	6	0
24/01/2012	14	32	8	17	4	0
25/01/2012	19	26	7	13	6	4
10/02/2012	20	28	10	6	11	0
05/03/2012	16	35	10	4	10	0
06/03/2012	19	28	12	10	6	0
09/03/2012	15	35	14	7	4	0
10/03/2012	13	35	11	9	6	1
11/03/2012	19	28	10	7	10	1
12/03/2012	17	33	12	9	4	0

For the Monte Carlo based methods, repeated trials are required. For clearer graphic presentation, only ten of the modelled load profiles were compared against the recorded daily aggregated load for one of the test days and the results are shown in Figure 5.20. Histogram of the load profiles showing their distributions are shown in Figure 5.21 with the first subplot showing the empirical one. It can be seen that the reconstructed load profiles fit well with the empirical profile. Figure 5.22 (a) - (b) uses whisker plots to show the reconstructed loads of the ten runs and of one hundred runs, respectively.

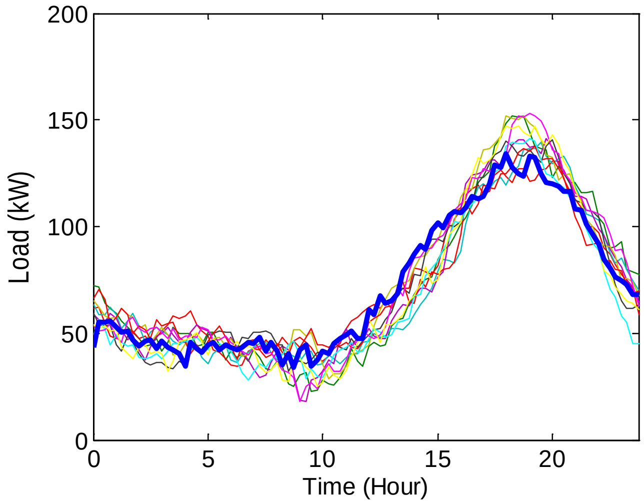


Figure 5.20 Ten reconstructed load profile for one of the high demand test days with the thicker curve showing the empirical load.

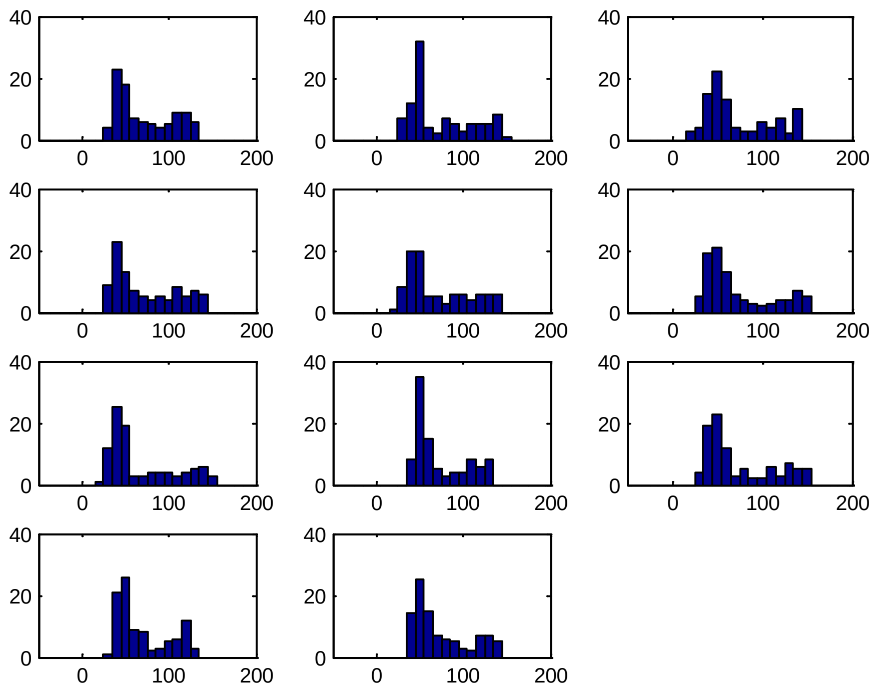
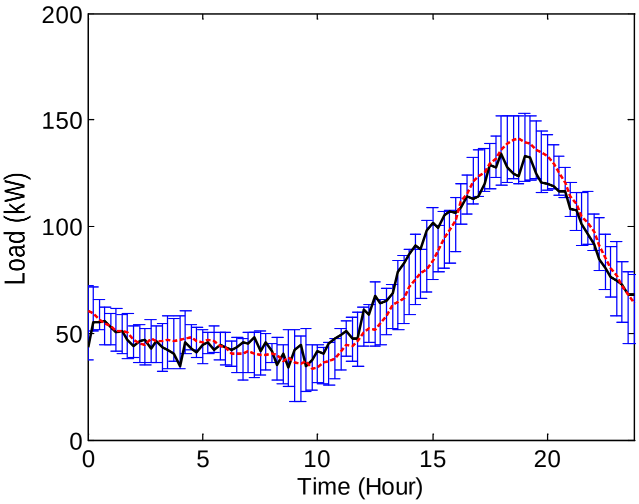
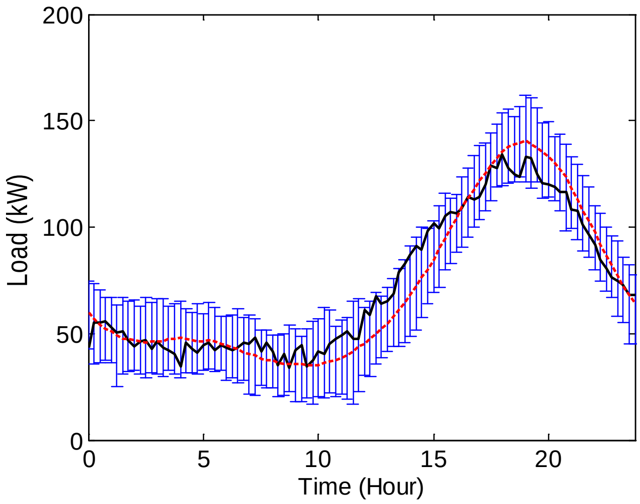


Figure 5.21 Histograms of the above 11 load curves, with the first one formed by the empirical load.



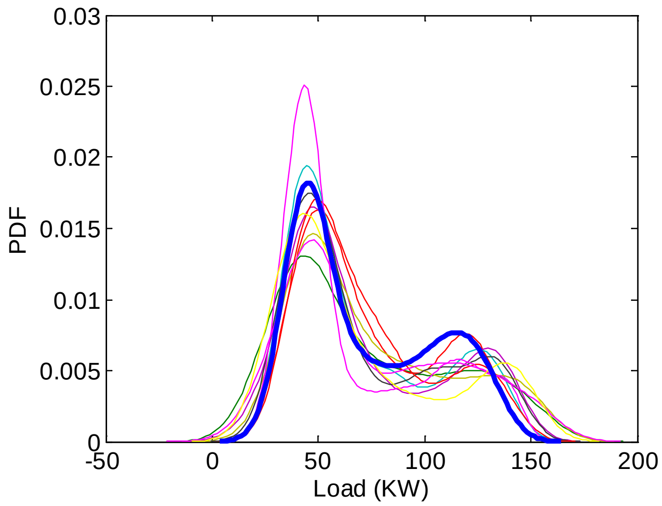
(a)



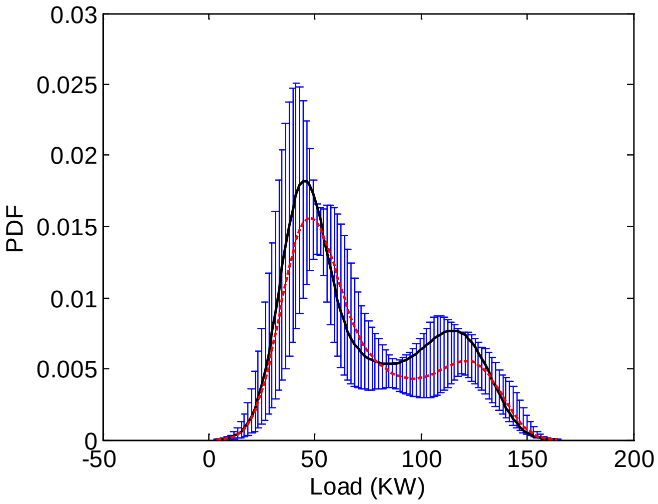
(b)

Figure 5.22 Whisker plots showing the ten and a hundred reconstructed load profiles in the high demand load model, respectively. The solid lines show the empirical load and the dashed lines show the mean of the reconstructed loads.

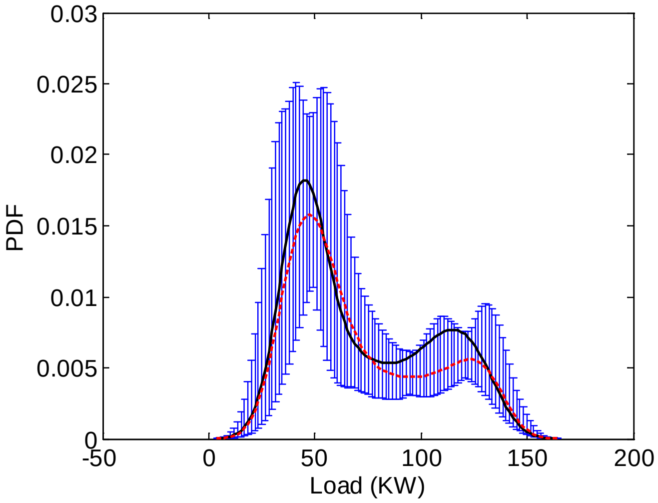
The fitness of the load reconstruction model could be evaluated by probability density function (PDF) and cumulative probability distribution function (CPDF). Figure 5.23 shows the PDF curves and the whisker plots of ten and a hundred reconstructed loads on one of the test days. Figure 5.24 shows the CPDF curves and the whisker plots of ten and a hundred reconstructed loads, respectively. The results clearly show that an ensemble of individual house loads, with the same cluster memberships as a specific test day, produces load profiles, PDFs and CPDFs that compare favourably with the physical daily load data. On all the five test days, reconstructed load profiles, PDFs and CPDFs were compared with the physical daily aggregated load data as shown in Figure 5.25 - Figure 5.27, and the curves all fits well with the empirical load data.



(a)

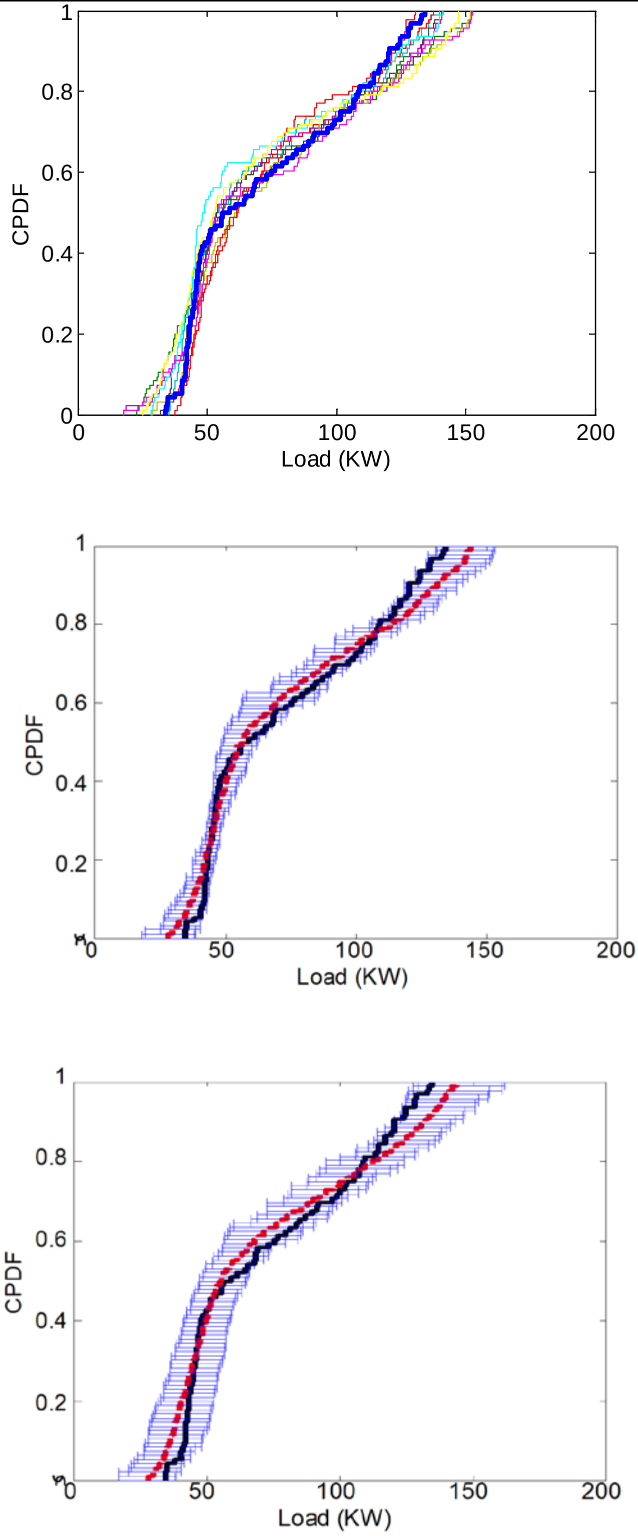


(b)



(c)

Figure 5.23 (a) PDFs of the ten reconstructed loads in the high demand load model with the thicker curve showing the empirical one. (b) – (c) Whisker plots of PDFs of the ten and a hundred reconstructed loads, respectively, with the solid line showing the empirical one and the dashed line showing the mean.



(a)

(b)

(c)

Figure 5.24 (a) CPDFs of the ten reconstructed loads in the high demand load model with the thicker curve showing the empirical one. (b) – (c) Whisker plots of CPDFs of the ten and a hundred reconstructed loads, respectively, with the solid line showing the empirical one and the dashed line showing the mean.

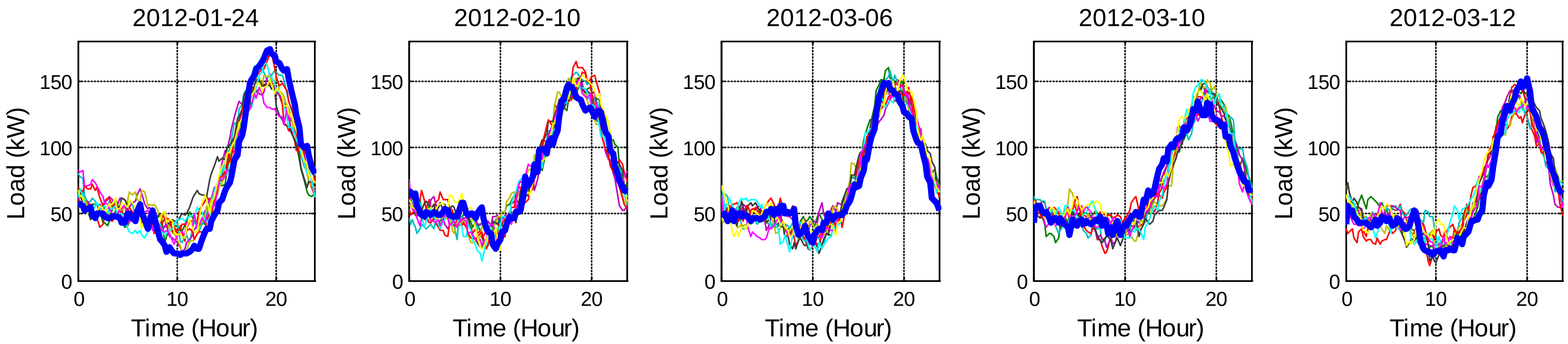


Figure 5.25 Ten reconstructed load profiles in the high demand load model with the thicker curve showing the physical load, each subplot showing profiles on each of the five test days, respectively.

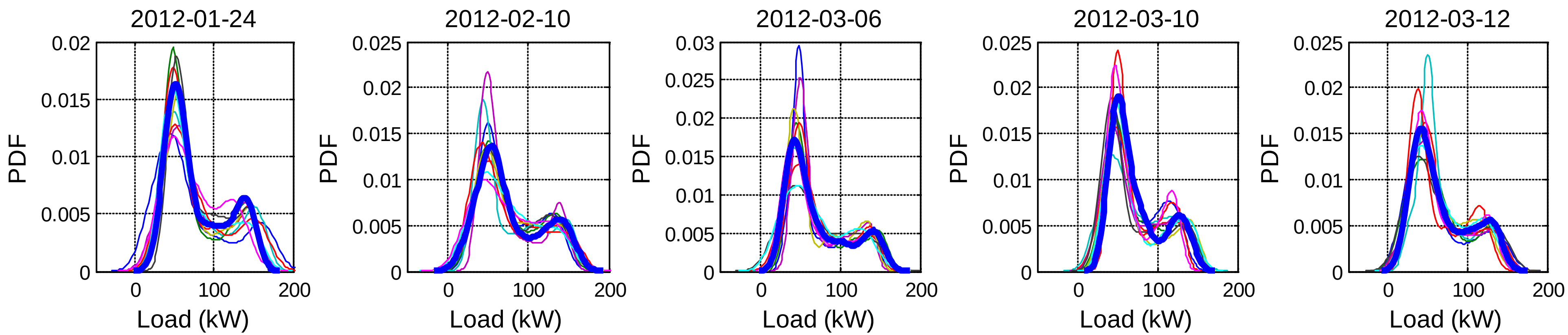


Figure 5.26 PDF curves of the ten reconstructed load profiles in the high demand load model with the thicker curve showing the physical load, subplots showing profiles on the five test days, respectively.

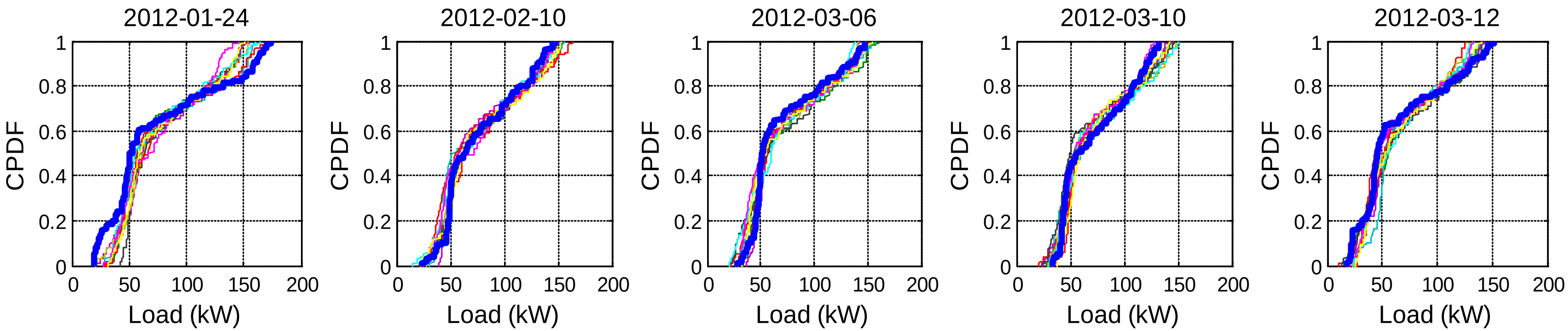


Figure 5.27 CPDF curves of the ten reconstructed load profiles with the thicker curve showing the physical load, subplots showing profiles on the five test days, respectively.

5.7 Summary

This chapter presents a method to develop a hybrid model for hot-day modelling of consumer load. A model for underlying diurnal variation is constructed using a truncated Fourier series representation. Cluster analysis is used to extract six representative daily load profiles. For each profile the variation in the Fourier coefficients was modelled using a normal distribution. After the low frequency variation is established a high frequency residual remains. This is modelled using a power spectrum. For Monte Carlo applications, models for a given consumer, within a cluster class, are established by:

- Selecting a set of Fourier coefficients according to the coefficient distributions for the selected cluster;
- Establishing a diurnal variation using the Inverse DFT;
- Establishing the power spectrum for the high frequency residual by selecting the magnitudes from the average magnitude profile for that cluster, and by adding a random phase selected from a uniform distribution;
- Producing the time series high frequency residual by applying the inverse transform;
- Combing the two to produce the final load model.

The models were validated using the data from the five high demand test days. For each test day ten load ensembles were constructed that included appropriate numbers of residential loads of each cluster type. The load groups successfully produced aggregated load behaviours that compare well with the physical load data in terms of load profile, PDFs and CPDFs.

References:

[1] D. H. O. McQueen, P. R. Hyland, and S. J. Watson, 'Monte Carlo simulation of residential electricity demand for forecasting maximum demand on distribution

networks', *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1685-1689, 2004.

[2] Y. Li, and P. Wolfs, 'Statistical Identification of Prototypical Low Voltage Distribution Feeders in Western Australia', in Power and Energy Society General Meeting, IEEE, 2012, pp. 1-8.

[3] F. Demailly, O. Ninet, and A. Even, 'Numerical tools and models for Monte Carlo studies of the influence on embedded generation on voltage limits in LV grids', *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp. 2343-2350, 2005.

[4] R. Singh, B. C. Pal, and R. A. Jabr, 'Statistical Representation of Distribution System Loads Using Gaussian Mixture Model', *Power Systems, IEEE Transactions on*, vol. 25, no. 1, pp. 29-37, 2010.

[5] P. Cugnet, 'Confidence Interval Estimation for Distribution Systems Power Consumption by using the Bootstrap Method', Electrical Engineering, Virginia polytechnic institute and state university, Blacksburg, Virginia, 1997.

[6] M. Davies, and R. G. Paterson, 'Phase unbalance in low-voltage distribution systems', *Proceedings of the IEE - Part A: Power Engineering*, vol. 109, no. 48, pp. 535-541, 1962.

[7] W. L. Briggs, and V. E. Henson, *The DFT: An Owners' Manual for the Discrete Fourier Transform*: Society for Industrial and Applied Mathematics 1987 pp. 16-25.

[8] G. W. Milligan, and M. C. Cooper, 'An Examination of Procedures for Determining the Number of Clusters in a Data Set', *Psychometrika*, vol. 50, pp. 159-179, 1985.

[9] M. C. Cooper, and G. W. Milligan, 'The Effect of Error on Determining the Number of Clusters', in Proceedings of the International Workshop on Data Analysis, Decision Support and Expert Knowledge Representation in Marketing and Related Areas of Research, 1988, pp. 319-328.

[10] *SAS/STAT® 9.2 User's Guide*, SAS Institute Inc. Cary, NC, USA, 2008.

[11] Y. Li, and P. Wolfs, 'Statistical discriminant analysis of high voltage feeders in Western Australia distribution networks', in Power and Energy Society General Meeting, IEEE, 2011, pp. 1-8.

Chapter 6 Application of load model and verification on low demand days

In this chapter the models and its verification for low demand days were developed as a comparison to the study on high demand days in the previous chapter.

The photovoltaic trial is subjected to local weather conditions. For a city like Perth with warm winters and hot summers, the residential daily load usually reaches its maximum on hot summer days, as residential consumers tend to use air-conditioners more on days when daily maximum temperatures are high. Demand data on the hottest days, picked according to its daily maximum temperature, were used in load modelling in Chapter 5. As a comparison, it would be interesting to build and test the models on the days with the possible lowest demand. On these days, the demand is negative and we observed reverse power flow at the middle of the day.

6.1 Model for low demand days

6.1.1 The selection of data for low demand days

The Pavetta Street transformer is fitted with a logger and the demand had been recorded on each of three phases independently in five-minute intervals during the PV trial. The sum of three values gives the demand from the transformer side. The ten days were picked when the demands from the transformer side were the lowest, and the lowest consumption and its time point on each day is listed in Table 6.1.

The meteorological data by Australian Bureau of Meteorology for Station Number 9021 (Airport) was used, because this station is closest to Pavetta Street. Table 6.1 also listed the meteorological information relevant to electricity consumption/generation on the ten days on daily global solar exposure, sun-hours, rainfall amount, evaporation,

TABLE 6.1
Summary of Meteorological Information on the Ten Days with High Solar Exposure

Date	**Time of minimum consumption**	**Value of lowest energy consumption (W)**	**Daily global solar exposure (MJ/m^2)**	**Sun-hours (H)**	**Rainfall amount (mm)**	**Evaporation (mm)**	**Maximum temperature (°C)**	**Minimum temperature (°C)**
8/11/2011	12:30	-11360	26.7	10.6	13.6	07.2	20.8	15.3
9/11/2011	12:00	-10090	28.1	10.6	01.6	05.2	21.8	10.5
10/11/2011	12:20	-11430	30.3	10.4	00.0	04.6	23.3	09.1
11/11/2011	11:30	-14980	30.7	11.5	00.0	05.6	25.4	10.1
14/11/2011	11:25	-10620	26.5	07.5	00.0	05.6	24.1	18.3
16/11/2011	11:05	-10830	31.8	12.1	00.0	07.4	23.9	09.4
28/11/2011	11:55	-15600	29.7	N/A	00.0	07.8	22.6	13.2
29/11/2011	11:35	-11190	33.4	13.4	01.4	06.8	25.0	12.7
30/11/2011	11:10	-15280	34.0	12.7	00.0	10.4	28.2	12.4
13/12/2011	13:40	-14090	21.3	04.4	35.2	10.2	24.2	17.4

maximum temperature, minimum temperature. Five validation days marked with shading in Table 6.2 (i.e. 22nd November 2011, 29th November 2011, 2nd December 2011, 10th December 2011and 17th December 2011) were used as test cases, while the other five days (i.e. 23rd November 2011, 30th November 2011, 3nd December 2011, 16th December 2011and 26th December 2011) were used to build the load model.

6.1.2 Load model for low demand days

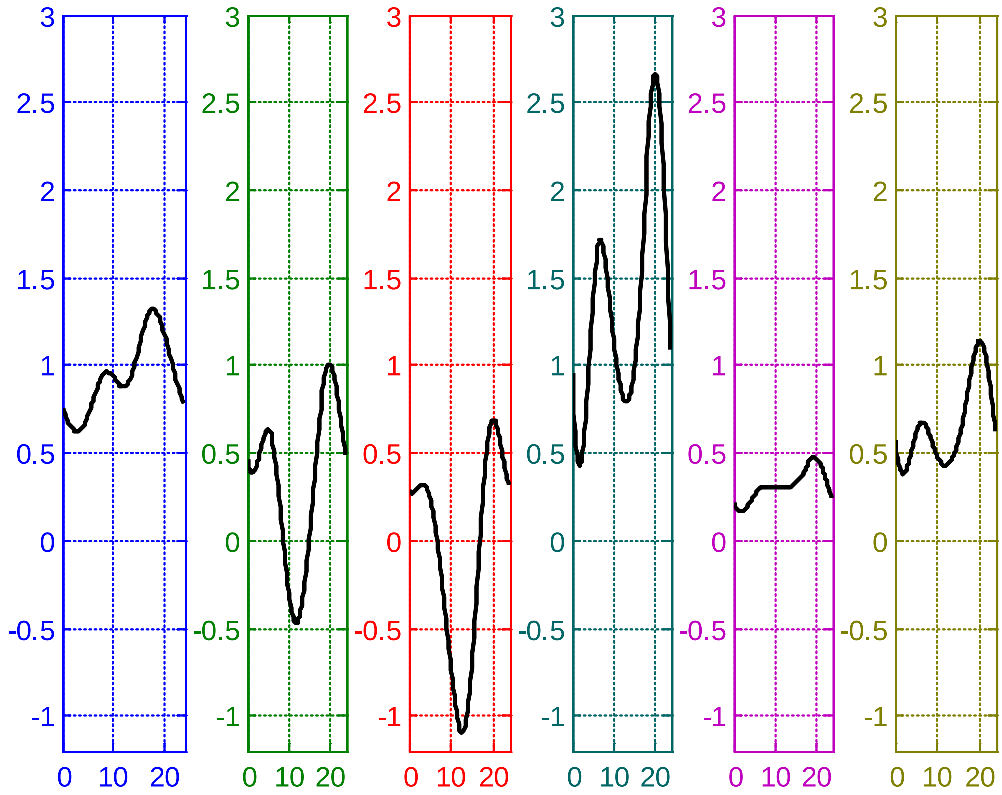
Similar to the modelling method introduced in Chapter 5, the first step is to combing the selected load data on the five days for building model into one data set to produce a large enough data set for clustering analysis. Then a finite sequence of real or complex numbers of the DFT were extracted. Similar to the selecting method illustrated in Figure 5.7 in Chapter 5, the vector length of four was identified as the optimal number. After DFT and extracting values for A_0, A_1, A_2, A_3, B_1, B_2 and B_3, clustering process was carried out using a combination of hierarchical and discriminant cluster analysis.

The optimal number of clusters was analysed by the three statistical parameters of Semi-Partial R^2, pseudo-F statistic and pseudo-T^2 Test, and six was determined to be the optimal number of clusters. The hierarchical cluster analysis was followed by several rounds of discriminant clustering analysis until a stable clustering and its cluster classifiers were obtained.

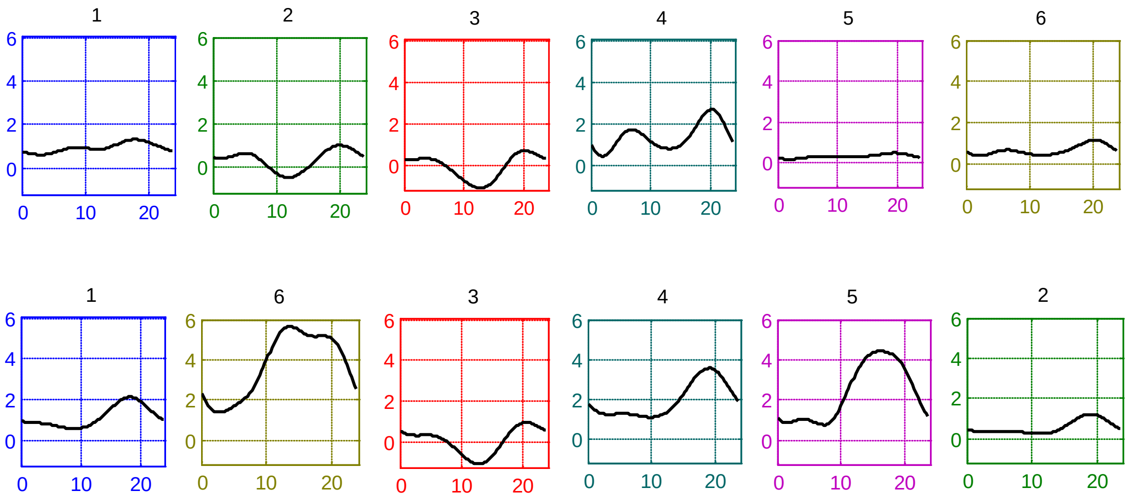
6.1.2.1 Load model on the periodic components

When constructing a load profile for a Monte Carlo simulation the Fourier coefficients are selected by generating values using the normal distribution parameters. Figure 6.1 (a) shows the low frequency periodic components for each of the six prototypical loads identified in the Pavetta Street feeder on low demand days. Figure 6.2 shows the modeled low frequency components of load for each of the 75 customers on Pavetta Street on a specific modelling day.

Cluster one, 51 of the 375 profiles (13.6%), is a customer with higher demand. Cluster two, 55 of the 375 profiles (14.7%), contains consumers with lower energy demands.

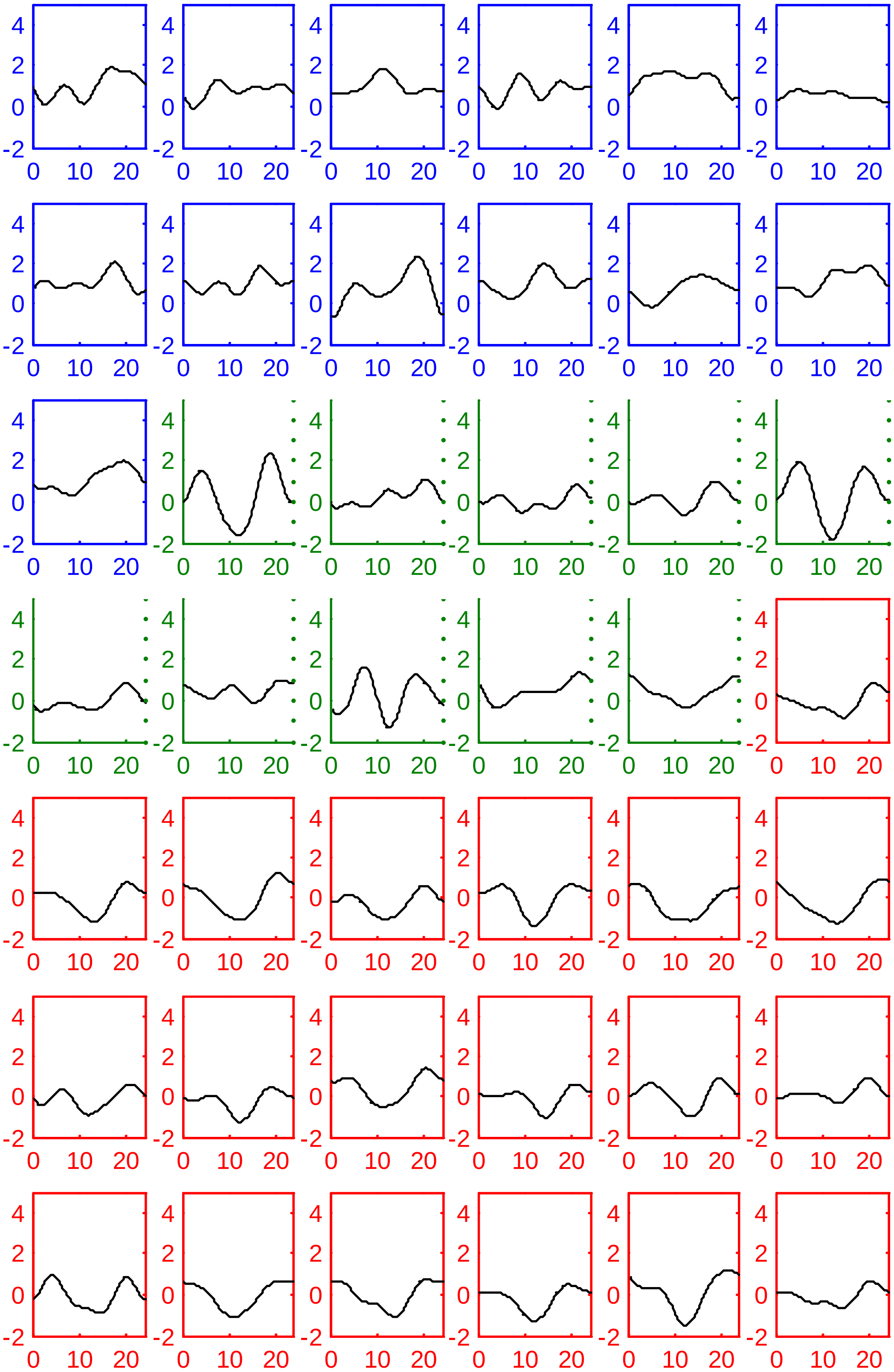


(a)



(b)

Figure 6.1 (a) The periodic components of the six prototypical loads on Pavetta Street on low demand days (x label: Time (Hour); y label: Load (kW)). (b) Comparison between low frequency periodic components for the six prototypical loads on low (upper figure) and high (lower figure) demand days.



[cont.]

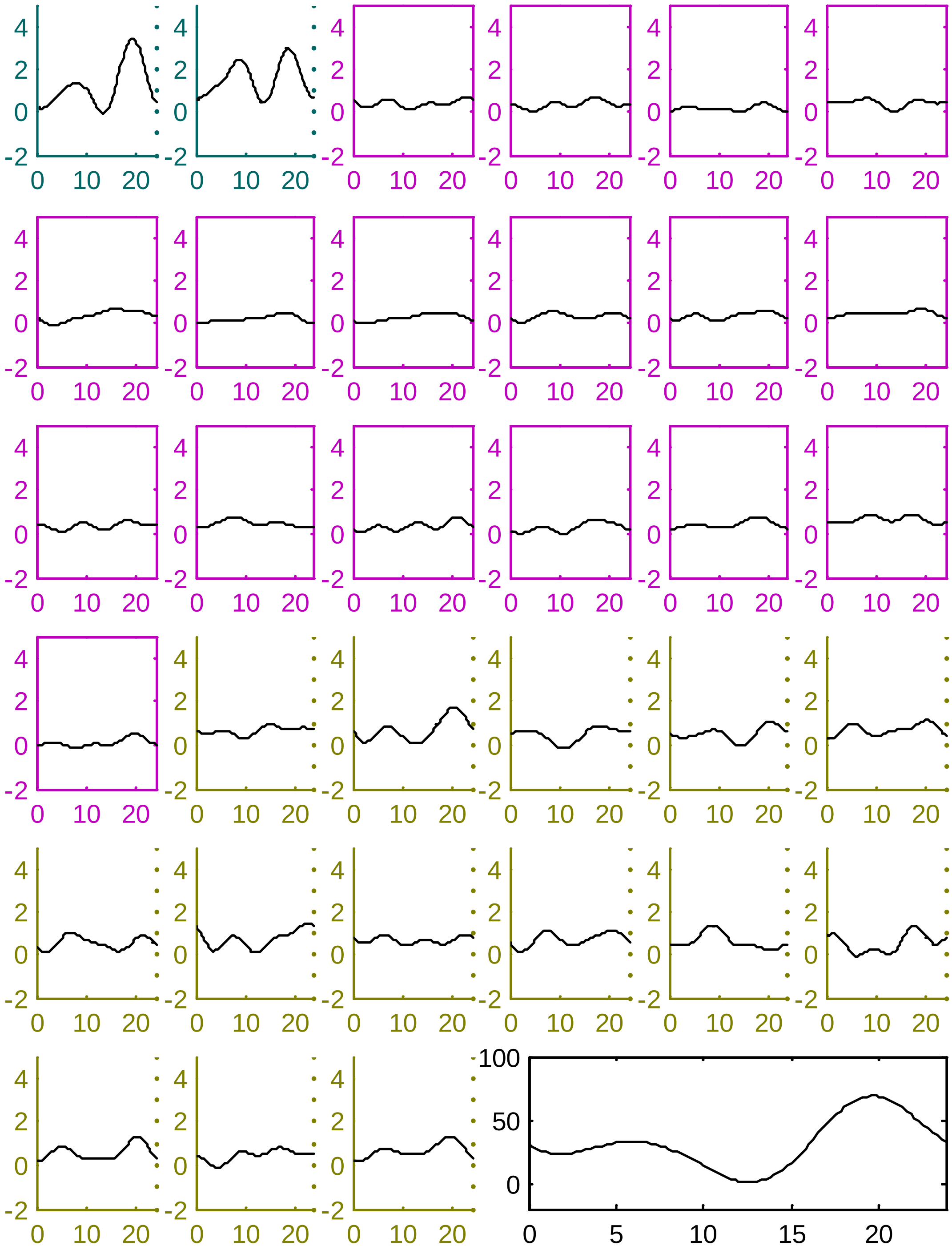


Figure 6.2 Modeled low frequency components of load for each of the 75 customers on Pavetta Street on low demand days. The subplot boxes of cluster 1, 3, 5 use solid lines, and 2, 4, 6 use dashed lines. The aggregated curve is shown at the end subplot.

Cluster three, shows the highest penetration of domestic roof top solar systems. This represented 88 of 375 (23.5%). Cluster four, shows consumers with the highest energy consumption and 13 of 375 (3.5%) profiles fell into this category. Clusters five and six, with 103 (27.5%) and 65 (17.3%) members respectively, are moderate energy consumers. Cluster five is the largest cluster.

The low frequency periodic components for the six prototypical loads on low and high demand days were put on the same scale in Figure 6.1 (b) in order to make a comparison between the two cases. It can be seen there are differences in the peak demands in the periodic components of the prototypical loads. In general, the maximum of peak demands nearly halved relative to heavy days. The two demand peaks are more obvious on the low demand days. With the increase of PV generation in the low demand days study, the demand dips more compared to the high demand day case study. The comparison shows the low consumptions on electricity were primarily due to the increase in PV generation and little air-conditioning usage. Also, the 3rd periodic component is nearly the same for the low and high demand day models.

6.1.2.2 Load model for the high frequency components

The difference between the physical and the low frequency Fourier model for consumer load is the high frequency component. An averaged power spectrum was calculated for each of the six load profile clusters and is seen as Figure 6.3. Figure 6.4 shows the modeled high frequency components of load for each of the 75 customers on Pavetta Street.

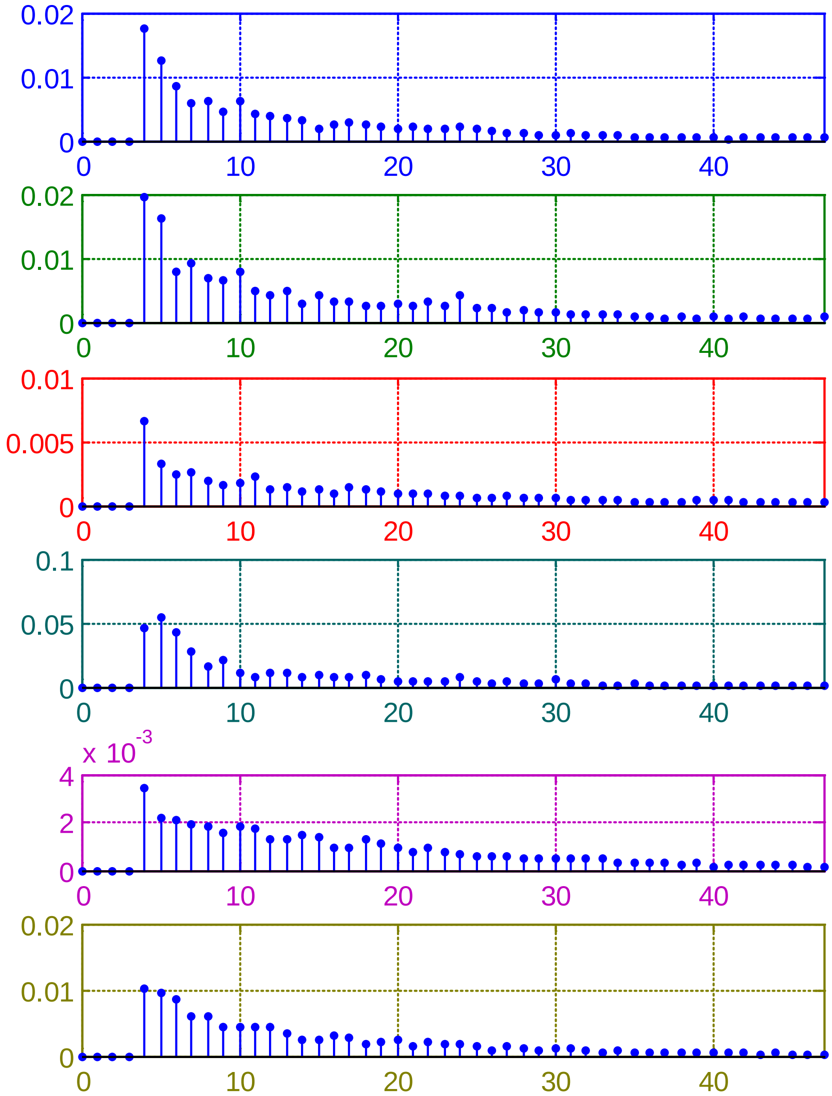
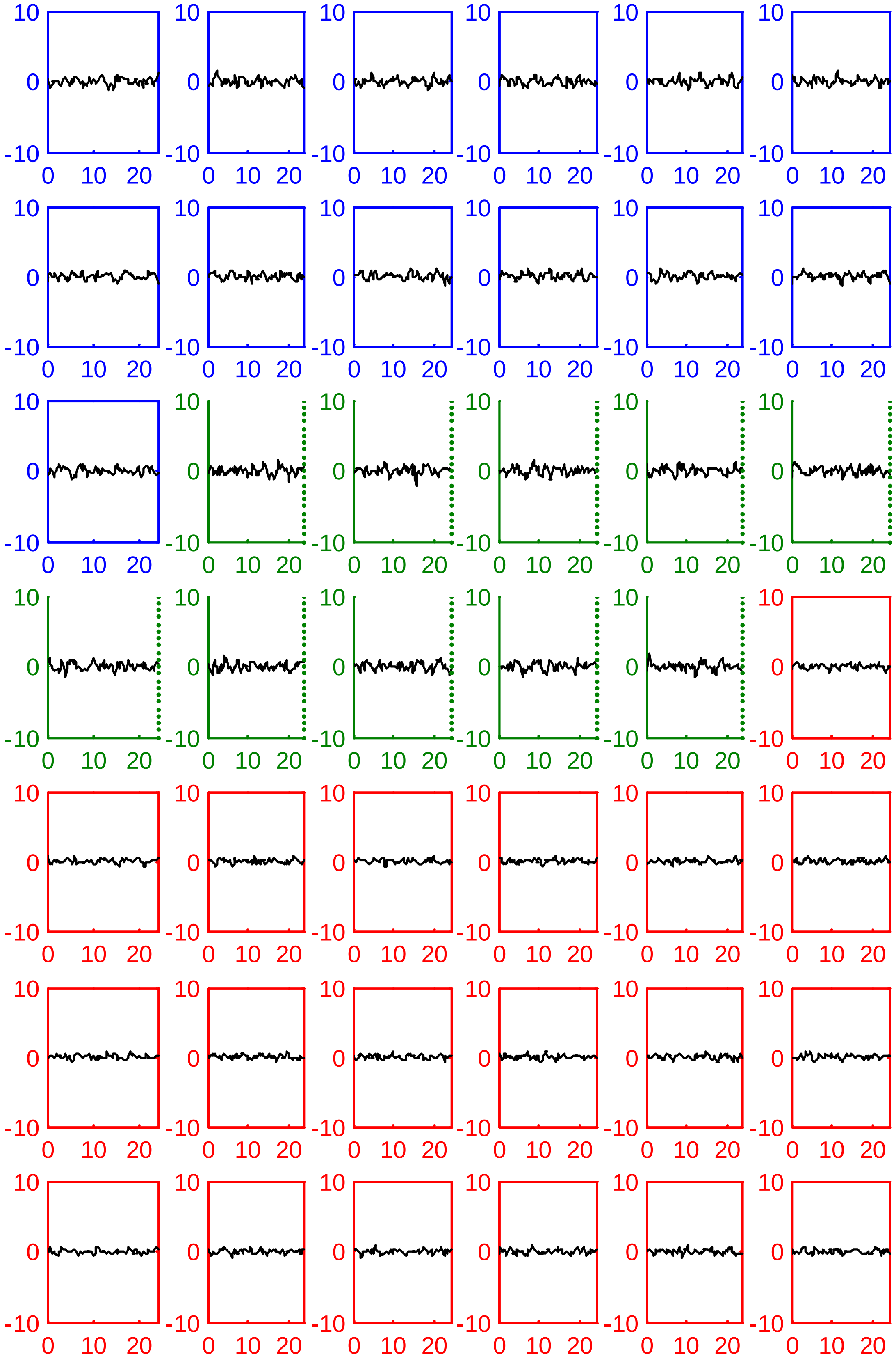


Figure 6.3 Frequency spectrogram for the six prototypical load profiles on low demand days. (x label: Frequency; y label: Averaged power spectrum)



[cont.]

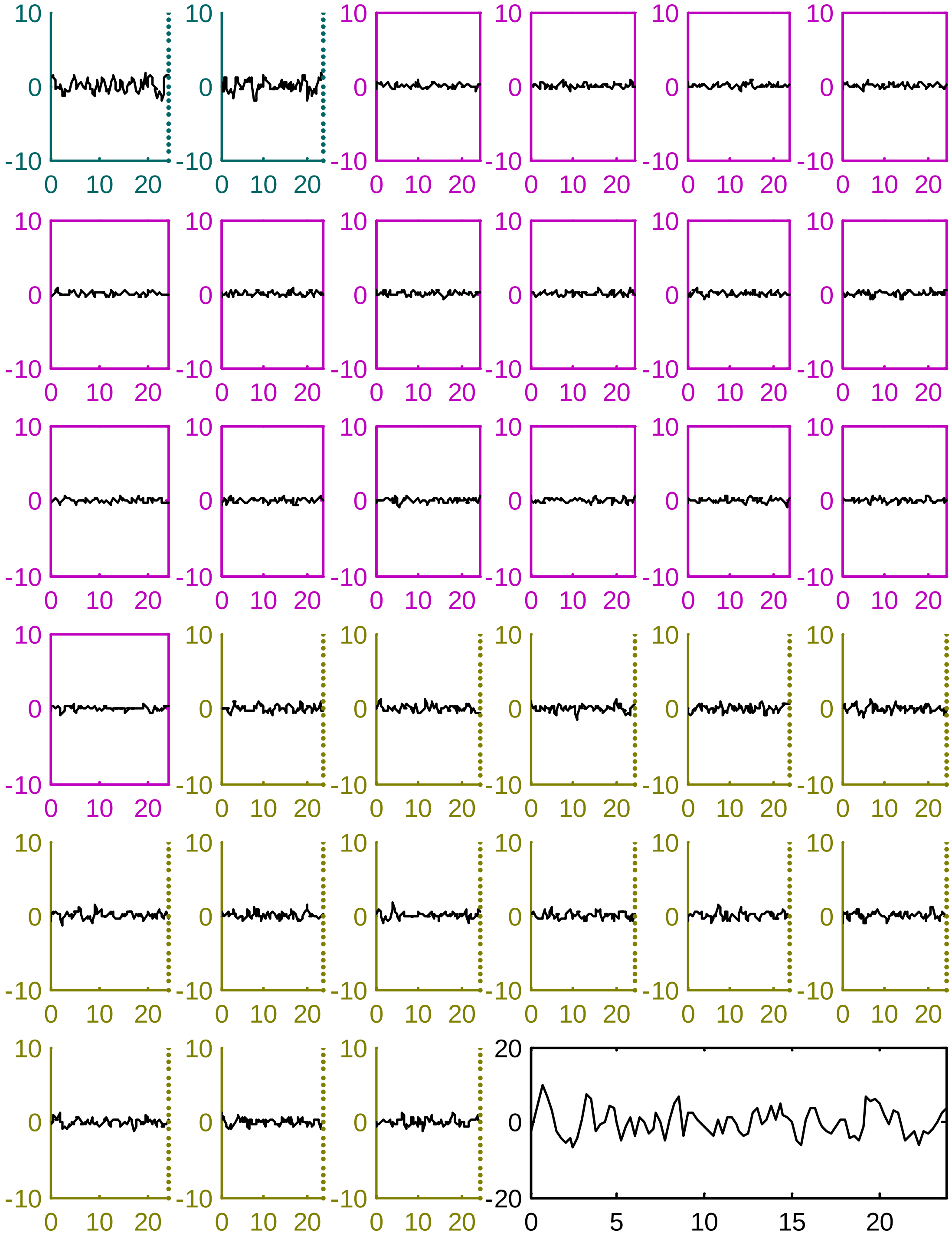
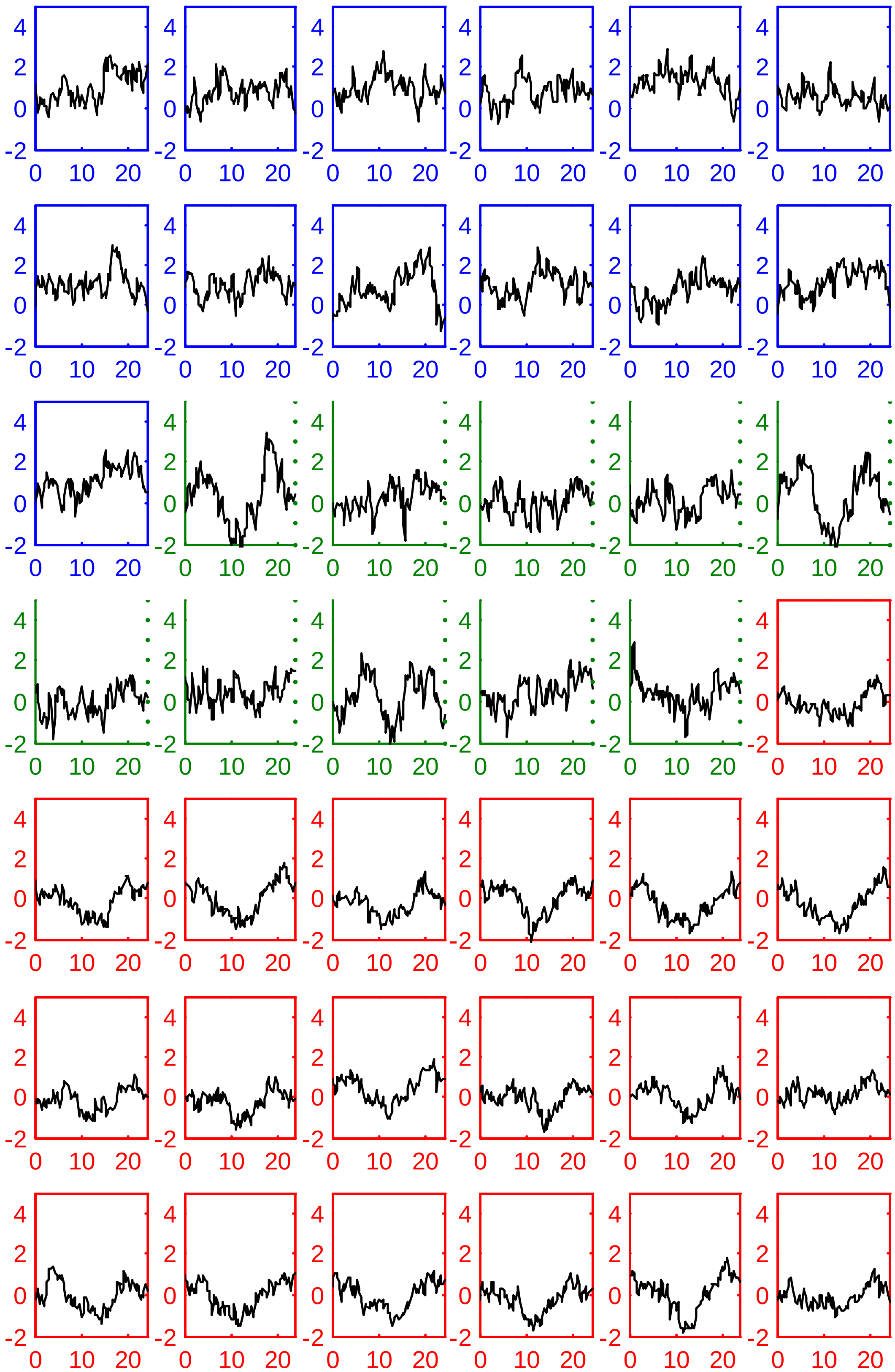


Figure 6.4 The modeled high frequency components of load for each of the 75 customers on Pavetta Street on low demand days.

The subplot boxes of cluster 1, 3, 5 use solid lines, and 2, 4, 6 use dashed lines. The aggregated curve is shown at the end subplot.

6.1.2.3 Load model results



[cont.]

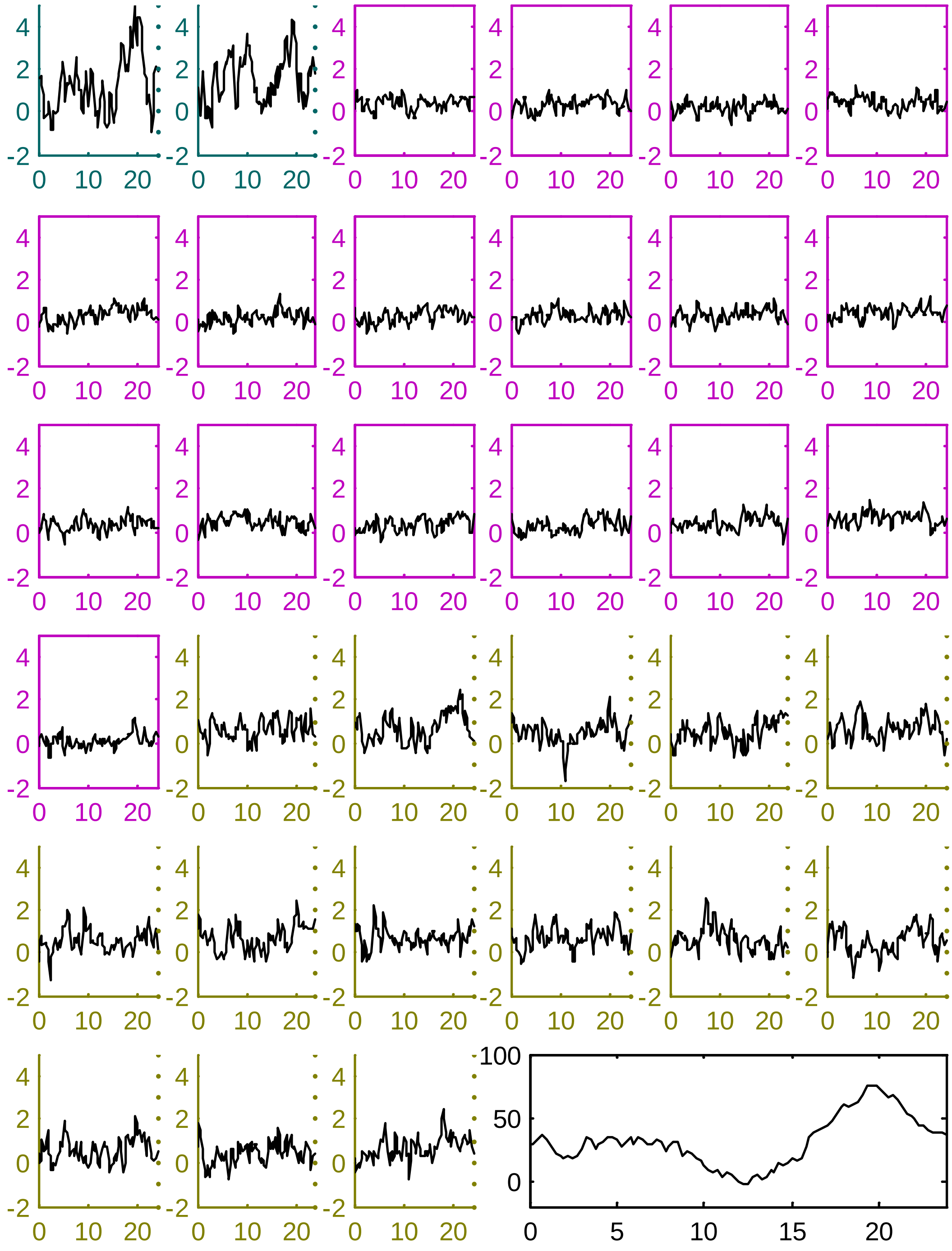


Figure 6.5 The modeled low and high frequency components of load for each of the 75 customers on Pavetta Street on low demand days. The subplot boxes of cluster 1, 3, 5 use solid lines, and 2, 4, 6 use dashed lines.

An example of a final reconstructed load profile can be built by combining the low and high frequency components as seen in Figure 6.5. Its last subplot shows the modeled aggregated daily load. Figure 6.6 shows the reconstructed load profiles with the thicker curve showing the empirical load. Figure 6.7 shows whisker plots of the ten and a hundred reconstructed load profiles, respectively.

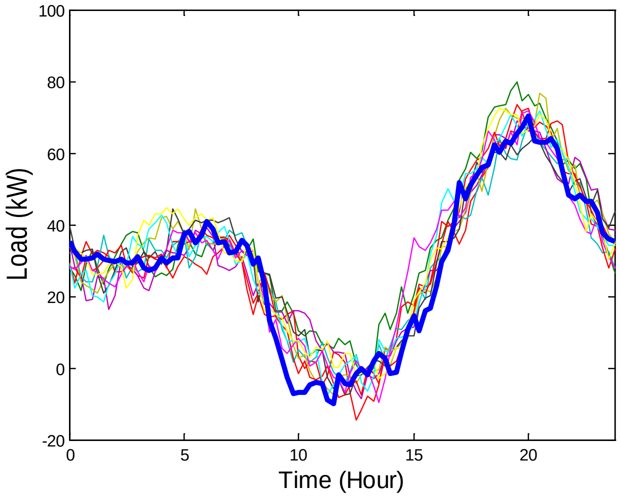
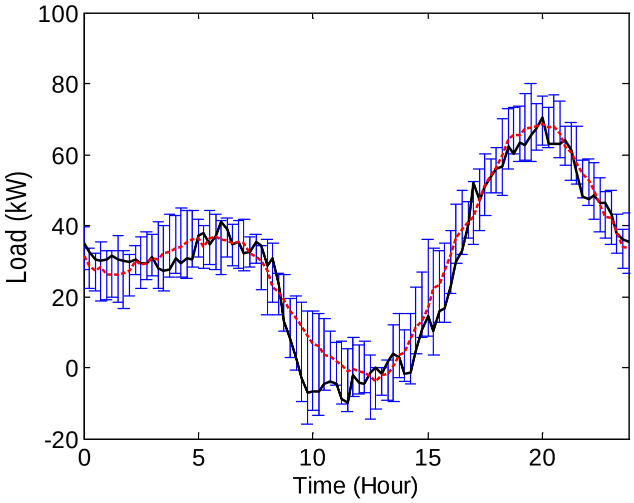
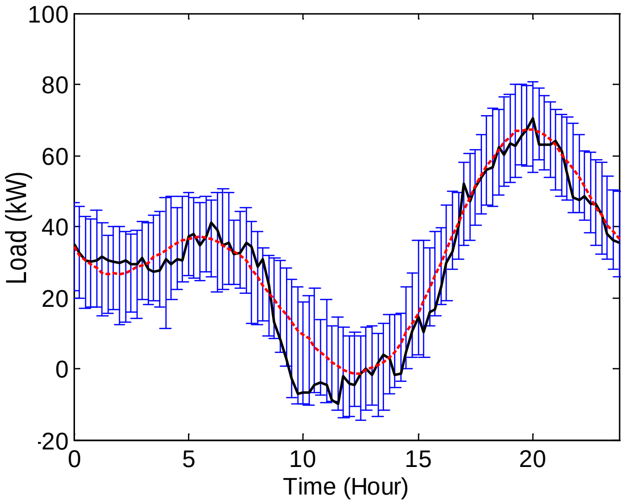


Figure 6.6 Ten reconstructed load profiles on low demand days with the thicker curve showing the empirical load.



(a)

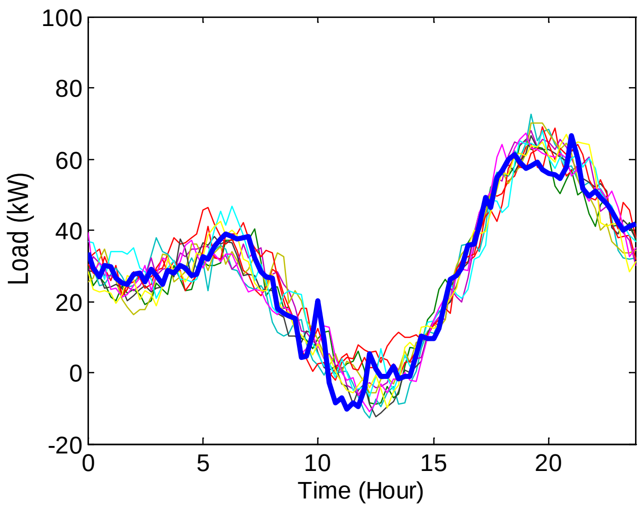


(b)

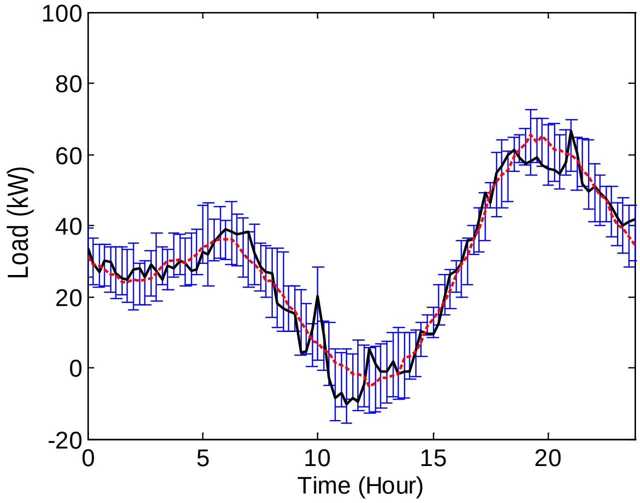
Figure 6.7 Whisker plots showing the ten and a hundred reconstructed load profiles on low demand days. The solid lines show the empirical load and the dashed lines show the mean of the reconstructed loads.

6.2 Load verification for low demand days

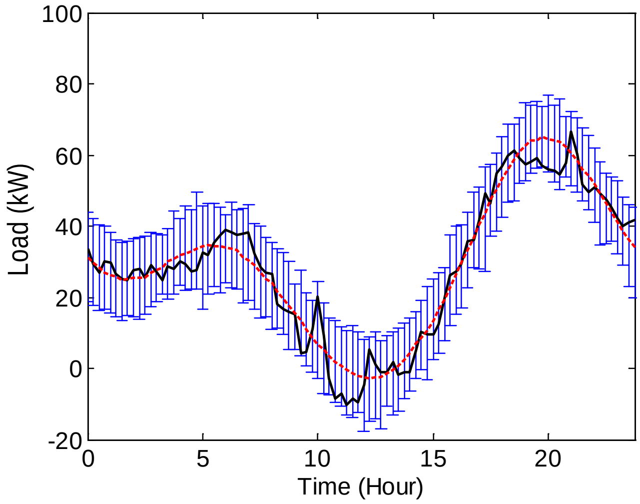
For one of the test days, ten of the modelled load profiles were compared against the recorded daily aggregated load as shown in Fig. 6.8 (a). It can be seen that the reconstructed load profiles fit well with the empirical profile. Figure 6.8 (b) - (c) uses whisker plots to show the reconstructed loads of the ten runs and of one hundred runs, respectively.



(a)



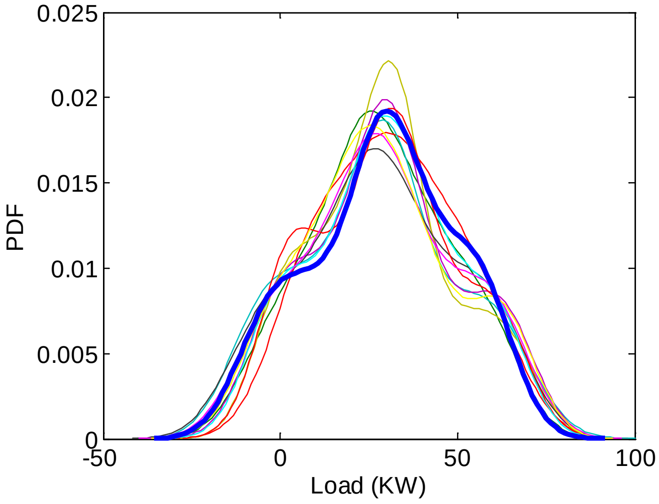
(b)



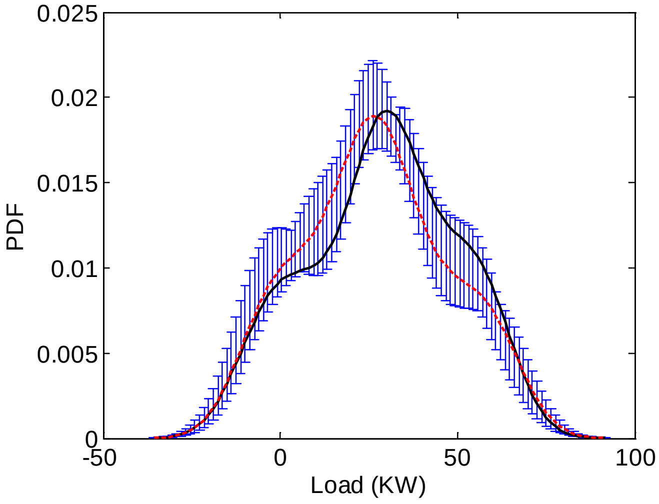
(c)

Figure 6.8 (a) Ten of the reconstructed load profiles with the thicker curve showing the physical load, each subplot showing profiles on each of the five low demand test days, respectively. (b) - (c) Whisker plots showing the ten and a hundred reconstructed load profiles, respectively. The solid lines show the empirical load and the dashed lines show the mean of the reconstructed loads.

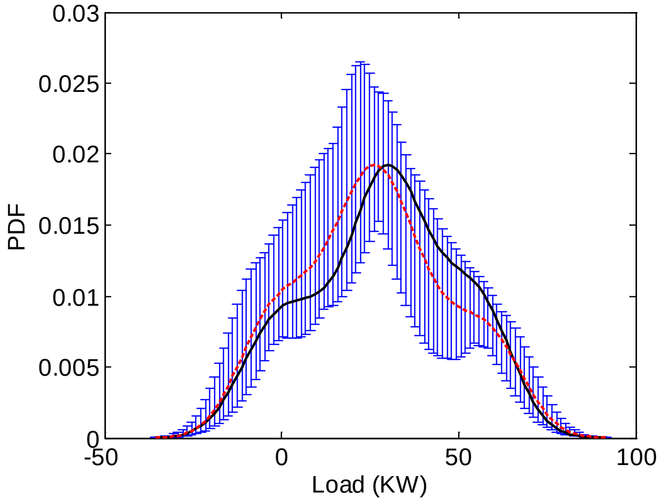
The fitness of the load reconstruction models were evaluated by PDFs and CPDFs, as well. Figure 6.9 shows the PDFs and the whisker plots of ten and a hundred reconstructed loads on one of the test days. Figure 6.10 shows the CPDFs and the whisker plots of ten and a hundred reconstructed loads, respectively. The results clearly show that an ensemble of individual house loads, with the same cluster memberships as a specific test day, produces load profiles, PDFs and CPDFs that compare favourably with the physical daily load data.



(a)

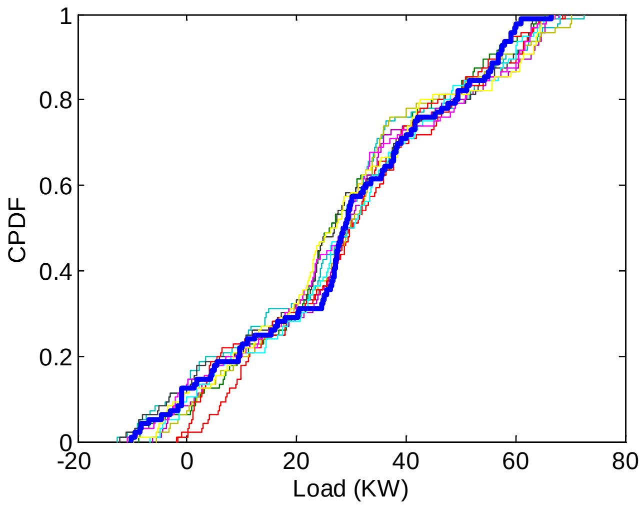


(b)

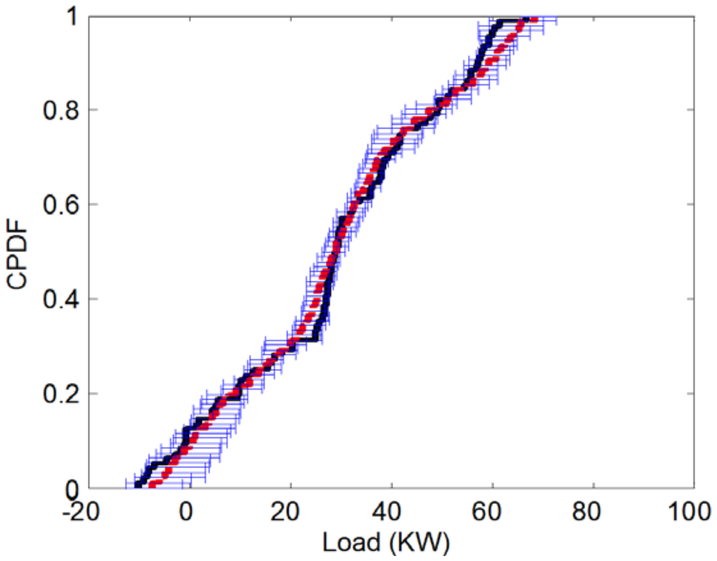


(c)

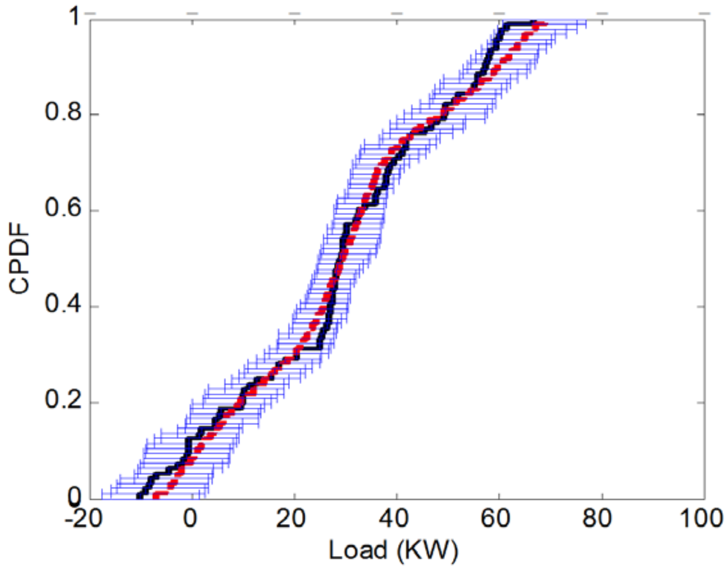
Figure 6.9 PDFs of the ten reconstructed loads on low demand days with the thicker curve showing the empirical one. (b) – (c) Whisker plots of PDFs of the ten and a hundred reconstructed loads, respectively, with the solid line showing the empirical one and the dashed line showing the mean.



(a)



(b)



(c)

Figure 6.10 CPDF of the ten reconstructed loads with the thicker curve showing the empirical one. (b) – (c) Whisker plots of PDFs of the ten and a hundred reconstructed loads, respectively, with the solid line showing the empirical one and the dashed line showing the mean.

6.3 Summary

In comparison to Chapter 5, models were applied to the case when the aggregated loads recorded from the transformer side were the lowest. The reconstructed loads fit well with the empirical aggregated load for the low demand days. The models were then validated using the data from the five test days. The load profiles, PDFs and CPDFs were found to compare well with the physical load data for the low demand days.

As it mentioned previously, the load model requires data sets at least at half hour intervals for it to be used in other case studies. During the PV trial, consumer load data have been recorded on 15 minute intervals. With finer recorded data, the accuracy level of modeled high frequency load components is expected to be increased.

Chapter 7 Conclusions and future work

7.1 Conclusions

This thesis aims to improve the fundamental understanding of the statistical and probabilistic nature of distribution feeders and their loads for network-wide assessment of Smart Grid technologies.

The thesis developed an efficient multivariable statistical analysis method that relies upon a few critical variables which are highly meaningful from an engineering perspective and readily available in most if not all distribution companies. Case-study based analyses resulted in the topological partitioning of MV and LV distribution network. The work identified statistically significant feeders using two very different data bases – a MV feeder data base of 204 members and an LV data base of 8,858 members. The capability of the general method to form useful cluster classifications in such disparate applications illustrates the universality of the method. The nine prototypical MV distribution feeders made use of the six selected variables. A set of discriminant functions has been extracted that can be used as the classifier for new MV feeders to be built in the WA distribution networks. Similarly, for the LV distribution feeders, eight prototypical LV feeders have been identified, making use of seven variables.

The thesis proposed a probabilistic load model aiming to capture the stochastic nature of residential loads and incorporating high penetration PV scenario. The hybrid models have been built and verified for high and low residential demand scenarios. After the low frequency diurnal variation, using a truncated Fourier series representation, was established, a high frequency residual remains. This has been modelled using a power spectrum. Cluster and discriminant analysis were used to extract six representative daily load profiles. The load models have been validated using the data from the test days. The load groups successfully produced aggregated load behaviours that compare well with the physical load data.

7.2 Research contributions

This thesis has made systematic study of improvements to distribution system modelling to support an array of emerging smart grid technologies. An impediment to the application of smart grid technologies is the development of the economic case for deployments. The scale and complexity of the distribution network makes it practically impossible to fully analyse the impact of a new technology in each and every corner of the network. A reasonable approach is to develop representative test cases, for both networks and consumer load to make the technology assessment process practicable. These representative test cases can then be applied in several ways. Some examples include:

- Representative network models for distribution feeders can be used with standard load flow tools to assess the impact upon voltage quality for new power electronic devices such as voltage regulators and distribution STATCOMS;
- Representative network models for distribution feeders can be used to assess the impact of conservation voltage reduction (CVR) on overall network energy and power demands [1];
- Representative network models for distribution feeders, and representative loads for consumer load, can be used with standard load flow tools to assess the voltage and load impact of distributed and community scale battery or fuel cell storages [2, 3];
- Representative network models for distribution feeders, and representative loads for consumer load, can be used with repeated load flow tools in a Monte Carlo sense to assess changes in the probability distributions of consumer voltage in the face of changes due to electric vehicle or solar generation uptake.

This thesis has made contributions in the following areas:

- Representative feeder models for both MV and LV distribution network;

- Representative models for consumer loads in the presence of high level of PV system uptake. These are now discussed in more detail.

Feeder Models

Australian distribution feeders are commonly classified as CBD, Urban, Rural Short and Rural or Rural Long, though it is not clear how the classifications have arisen. A rigorous assessment of any new technology first requires the establishment of test cases. Monte Carlo analysis is widely used to model distribution networks to study the impact of new load groups such as electric vehicles, or distributed generation such as roof top PV and demand management technologies. As the mathematics are based on probability theory and stochastic processes, valid or representative customer load data are needed as input, in order to extract useful results from this computationally intensive technique. The thesis makes contributions to knowledge under the following areas:

Issue - Studies have been published on the application of cluster analysis to MV feeders to extract representative or prototypical feeders, which has an excessive input data requirement – 37 variables per feeder in the PNNL case.

Contribution - This thesis developed a multivariate statistical method of higher utility by using fewer information rich variables that can be easily extracted from common DNSP data bases;

Issue - No published literature that comprehensively addresses statistical studies on LV feeder assessment.

Contribution - The thesis developed parameter selection approaches to assist in extracting representative feeder subsets from large LV feeder sets.

Feeder Classifiers

Issue - The feeder taxonomy studies have been carried out in literature, but none of these has given out classifiers for MV or LV feeders.

Contribution - The thesis applied cluster and discriminant analysis on MV and LV distribution feeders, identified prototypical feeders and given out feeder classifiers for each subsets by quadratic discriminant functions.

Load Demand Models – Diurnal Time Series

Issue - Modelling accuracy depends on the load group and the model.

Contribution - This thesis developed load models for residential customers with high PV penetration, which have not been statistically studied in the literature. The reconstructed daily load profiles, PDFs and CPDFs fits well with the empirical load data.

The outcomes of the thesis will assist the distribution network operators in their technical decision making and better understanding in the deployment of Smart Grid technologies. Assessment of Smart Grid technologies first requires the construction of test cases that are statistically meaningful. It makes it practical to test various Smart Grid technologies and models on a few typical feeders. Furthermore, the development of a more rigorous statistical model can be directly useful for any distribution operator in regard to informing their design processes. The use of the overlap of the Smart Meter deployment and the High Penetration PV Feeder project examined the agreement of residential load modelling and identified the impact of system voltages on loads when with and without PV generation. It allows studies to identify the power quality variation with consumer behaviour, and to look into the correlation of generation and load. Also, this project facilitates the uptake of renewable generation.

7.3 Future work

More work is underway out to produce models for the MV feeders in the IEEE test feeder format. After that, models for the LV feeders are expected to be developed in the IEEE test feeder format, too.

Hopefully, with further data availability in other states in Australia, the feeder taxonomy and load models could be built, and tested in different geographical territories other than Western Australia and in distribution networks with various new Smart Grid technologies.

During the PV trial, residential load data were recorded on 15 minute intervals. The accuracy level of high frequency components in the load models is expected to be increased with finer recoded data. The load models could be tested when finer data becomes available in the future.

Monte Carlo simulation based solutions are easy to formulate but are computationally intensive and may not provide an understanding of the underlying dynamics. A closed form analytic solution, if it can be found, provides a superior level of understanding and is capable of generalization. Future work would be carried out to determine the extent to which closed form or alternate numerical solution methods exist. It may be possible to develop analytical expressions for aggregated load expectancy and variation. For different load groups, a cross-correlation approach may allow quick estimates of aggregated behaviour to be made.

More work could be done aiming to develop a pathway for the integration of variable renewable generation into the power systems that minimises the capital investments by avoiding unnecessary network reinforcement while still retaining high power quality indices.

References:

[1] K. P. Schneider, J. C. Fuller, F. K. Tuffner, and R. Singh, *Evaluation of Conservation Voltage Reduction (CVR) on a National Level*, Pacific Northwest National Laboratory report, July 2010.

[2] B. C. H. Steele, and A. Heinzel, 'Materials for fuel-cell technologies', *Nature*, vol. 414, no. 6861, pp. 345-352, 2001.

[3] A. Khaligh, and L. Zhihao, 'Battery, Ultracapacitor, Fuel Cell, and Hybrid Energy Storage Systems for Electric, Hybrid Electric, Fuel Cell, and Plug-In Hybrid Electric Vehicles: State of the Art', *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 6, pp. 2806-2814, 2010.

APPENDIX: Coefficients used in quadratic discriminant functions

TABLE A.1

Coefficients Used in the Set of Quadratic Discriminant Functions for the example Set of MV Feeders in Chapter 3

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
$\mathbf{a_0}$	-102.394036	-150.907604	-33.554666	-40.223164	-26.436480	-964.500374	-63.441718	-74.070212	-64.059079
$\mathbf{a_1}$	6.227919	12.565677	-5.554895	1.582936	1.435782	-50.682964	1.658161	3.402431	3.023118
$\mathbf{a_{11}}$	-0.296506	-0.818233	-2.847078	-0.101546	-0.297764	-1.928721	-0.180423	-0.220869	-0.302292
$\mathbf{a_{12}}$	0.000033	0.000209	0.003789	0.000136	0.000039	0.056963	-0.000063	-0.000029	0.000469
$\mathbf{a_{13}}$	0.022334	-0.002035	0.003009	-0.010605	-0.000221	-0.470392	-0.000070	0.003818	-0.005406
$\mathbf{a_{14}}$	-0.008512	0.002963	0.094860	0.007157	-0.005844	0.075707	-0.000658	0.001542	0.007363
$\mathbf{a_{15}}$	-0.007390	-0.047356	0.262832	-0.004875	0.009111	1.467188	0.015399	0.007489	-0.013172
$\mathbf{a_{16}}$	0.005485	0.141637	0.190021	-0.003521	0.005042	0.208706	0.003557	-0.014067	0.029093
$\mathbf{a_2}$	-0.002459	0.017085	0.008120	-0.002602	-0.006451	2.016937	-0.000647	0.004322	-0.021185
$\mathbf{a_{22}}$	-0.000003	-0.000004	-0.000018	-0.000010	-0.000020	-0.002097	-0.000001	-0.000001	-0.000007
$\mathbf{a_{23}}$	0.000383	0.000075	0.000005	0.000543	0.000097	0.020280	0.000027	0.000032	0.000048
$\mathbf{a_{24}}$	0.000023	0.000050	0.000408	0.000139	0.000027	-0.002729	0.000006	0.000013	0.000015

[cont.]

a_{25}	0.000039	-0.000035	-0.001148	0.000041	0.000948	-0.050994	0.000070	0.000023	0.002199
a_{26}	-0.000172	-0.000579	-0.000426	-0.000038	-0.000033	-0.007886	-0.000066	-0.000296	-0.000097
a_{3}	1.994014	1.696768	0.037288	0.606275	0.144666	-16.023484	0.611879	0.173714	0.246530
a_{33}	-0.102381	-0.050280	-0.000038	-0.073593	-0.001053	-0.250527	-0.002971	-0.003550	-0.000456
a_{34}	0.007370	0.010146	-0.000249	0.006686	0.000008	0.023025	0.000398	0.000355	0.000207
a_{35}	-0.012223	-0.009843	-0.003134	-0.004636	-0.006314	0.354302	-0.006533	-0.002970	-0.019450
a_{36}	0.004338	0.007257	-0.000250	0.003310	-0.000421	0.067559	-0.008609	0.003892	0.000386
a_{4}	0.077243	-0.327320	-0.190984	-0.075176	0.005363	6.662401	-0.109272	-0.100740	-0.090188
a_{44}	-0.004288	-0.005186	-0.032099	-0.011539	-0.001097	-0.024920	-0.000573	-0.000861	-0.001483
a_{45}	0.001076	0.001669	0.101696	0.001577	0.000770	-0.181886	0.001800	0.000386	0.009613
a_{46}	0.001002	0.000815	0.005100	-0.000447	0.000571	-0.020030	0.004060	0.008080	-0.000975
a_{5}	0.890534	2.090207	8.298949	0.627932	0.821198	106.671425	1.373259	0.481041	12.528821
a_{55}	-0.003764	-0.008918	-1.554583	-0.005889	-0.063055	-3.222572	-0.021229	-0.004503	-1.349101
a_{56}	-0.010319	-0.010017	-0.096303	-0.006500	-0.001785	-0.330059	-0.023953	-0.011210	-0.018523
a_{6}	4.805318	4.778962	1.682807	1.075335	0.138086	12.095763	3.159755	6.573274	0.585562
a_{66}	-0.262640	-0.286018	-0.063866	-0.017683	-0.001575	-0.040653	-0.091388	-0.407832	-0.017637

TABLE A.2

Coefficients Used in the Set of Quadratic Discriminant Functions for MV Feeders

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
a_0	-1.4229E+02	-5.0346E+01	-1.0282E+02	-1.6983E+02	-1.6961E+02	-4.2889E+01	-9.6038E+01	-1.0588E+02	-3.6713E+08
a_1	6.5790E+00	1.2955E+00	3.1184E+00	2.0487E+01	-1.4756E+00	9.0397E-02	5.1548E+00	1.8843E+01	4.3790E+07
a_{11}	-4.0210E-01	-1.7473E-01	-1.7660E-01	-9.2120E-01	-3.4560E-01	-1.4729E-01	-1.3807E+00	-1.2412E+00	-1.3125E+06
a_{12}	-3.8400E-04	5.1992E-05	1.4050E-04	-8.8071E-04	3.8655E-04	1.7760E-03	4.2788E-03	-8.6260E-03	-2.0654E+03
a_{13}	3.6707E-02	5.7781E-03	-2.3996E-02	-1.7987E-02	3.1492E-02	-7.6226E-04	-3.1567E-02	9.0085E-02	-3.6965E+04
a_{14}	-3.8180E-03	-2.8174E-03	1.7939E-02	1.6642E-02	8.2495E-04	-3.2457E-03	-8.5293E-03	7.4371E-02	4.0685E+04
a_{15}	9.0133E-03	1.2248E-02	-7.4576E-03	-4.8974E-02	1.3950E-02	1.0108E-02	-5.4436E-02	2.8843E-01	-5.3236E+03
a_{16}	9.2282E-02	5.7154E-02	-2.1327E-02	-7.3830E-02	9.0179E-02	5.2508E-03	4.8755E-01	-7.9670E-01	-1.6306E+05
a_2	-1.5017E-03	-1.3287E-03	-4.3231E-03	2.0479E-02	6.6078E-03	-2.3590E-02	-1.6259E-02	1.5310E-01	6.7194E+04
a_{22}	-3.8456E-06	-1.2541E-06	-8.7120E-06	-6.0444E-06	-5.4033E-06	-7.6136E-05	-2.0791E-05	-1.4270E-04	-3.6441E+00
a_{23}	4.5659E-04	4.7422E-05	5.8352E-04	1.6210E-04	7.4170E-04	3.7243E-04	-4.5817E-05	1.4706E-03	-5.7218E+01
a_{24}	-4.3715E-06	7.2323E-06	1.0643E-04	1.7661E-05	-1.4739E-05	9.2597E-07	2.6881E-04	8.0302E-04	7.0313E+01
a_{25}	7.5537E-05	5.4701E-05	4.3105E-05	1.1183E-04	4.1161E-05	2.0411E-03	1.9298E-04	8.0070E-03	1.1042E+02
a_{26}	3.0525E-04	-1.5504E-04	-3.6634E-05	-2.2611E-04	-1.1679E-03	-2.1872E-05	-8.4638E-04	-1.0706E-02	-2.4384E+02
a_3	3.3422E+00	2.8249E-01	1.5424E+00	1.6684E+00	2.9800E+00	5.1654E-01	1.6314E+00	-1.4259E+00	1.2375E+06
a_{33}	-1.0470E-01	-4.1382E-03	-8.2338E-02	-2.5238E-02	-1.8218E-01	-5.1300E-03	-6.5801E-02	-1.5762E-02	-1.0434E+03
a_{34}	1.0574E-02	2.1608E-04	4.7281E-03	7.7582E-03	6.6254E-03	1.3951E-04	1.4306E-02	-8.2466E-03	1.1307E+03
a_{35}	-2.2872E-02	-4.1780E-03	-7.2944E-03	-1.6743E-02	-2.2302E-02	-2.4531E-02	-4.9659E-03	-8.5382E-02	-4.3722E+02

[cont.]

a_{36}	-7.5881E-02	-7.4071E-04	-7.7556E-03	-1.6444E-02	1.6276E-01	-1.1563E-03	6.2134E-04	1.1002E-01	-4.6234E+03
a_4	-1.6880E-01	7.5601E-02	-1.7151E-01	-6.5810E-01	3.1823E-01	5.2498E-02	3.6371E-01	-1.1965E+00	-1.3301E+06
a_{44}	-4.6949E-03	-6.0915E-04	-1.2177E-02	-5.5363E-03	-1.9282E-03	-7.1905E-04	-1.2727E-02	-6.0522E-03	-1.3616E+03
a_{45}	2.4233E-03	1.0822E-04	1.3211E-03	4.5340E-03	-1.3349E-03	3.4290E-04	-1.4201E-03	-3.5748E-02	-1.7310E+03
a_{46}	7.3411E-03	4.2670E-03	2.4485E-03	1.6905E-02	7.7239E-03	1.6700E-05	-3.3418E-02	6.6696E-02	4.8511E+03
a_5	1.7345E+00	4.5357E-01	2.0331E+00	2.1885E+00	2.4779E+00	3.9164E+00	1.2374E+00	-4.8652E+00	6.9534E+05
a_{55}	-1.0432E-02	-6.3723E-03	-1.4121E-02	-1.6243E-02	-1.1953E-02	-2.0313E-01	-6.5036E-03	-5.3938E-01	-3.5847E+04
a_{56}	-3.5084E-02	-7.1121E-03	-2.8579E-02	-1.5827E-02	3.1517E-02	-8.7795E-03	-8.1946E-03	5.5269E-01	-4.5101E+03
a_6	6.6303E+00	1.9078E+00	4.8201E+00	4.1221E+00	-1.1076E+00	2.8598E-01	8.5222E+00	1.4440E+01	5.4961E+06
a_{66}	-3.2135E-01	-1.7250E-01	-7.4879E-02	-1.5255E-01	-6.6357E-01	-1.0884E-03	-7.1110E-01	-8.6978E-01	-2.0672E+04

TABLE A.3

Coefficients Used in the Set of Quadratic Discriminant Functions for LV Feeders

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
a_0	-1.1041E+01	-2.4858E+01	-2.9472E+01	-1.1933E+01	-2.7381E+01	-6.7296E+00	-1.6332E+01	-1.2808E+01
a_1	1.8332E-01	7.2117E-01	2.7859E-01	3.5456E-01	7.0105E-02	3.5234E-02	6.1493E-01	3.4206E-01
a_{11}	-6.1818E-01	-2.2211E-01	-5.9134E-02	-1.8888E+00	-2.5641E-03	-9.5200E-03	-8.7411E+00	-2.2713E+00
a_{12}	-8.8166E-02	9.6475E-04	8.1258E-04	1.4775E-03	-9.1227E-05	-2.3228E-29	1.2561E-02	9.4380E-03
a_{13}	3.9014E-02	1.3175E-02	2.2135E-03	9.4266E-01	5.4797E-05	4.7918E-05	6.5343E+00	1.4853E+00
a_{14}	1.9081E-02	-1.2972E-03	-3.9229E-05	-1.4223E-04	-3.9320E-05	3.4175E-12	-6.7302E-03	-4.2308E-01
a_{15}	2.8552E-02	-1.4727E-01	-6.2226E-02	-2.2044E-01	-1.4812E-02	2.2820E-02	4.1202E-02	8.3861E+01
a_{16}	1.2009E-03	2.1630E-04	-4.6042E-05	4.1791E-11	5.9055E-06	2.0819E-04	6.9063E-04	1.4895E-04
a_{17}	9.1999E-03	-5.8050E-05	-1.2480E-05	1.7580E-04	1.6153E-05	6.5918E-05	-1.4419E-03	4.2539E-04
a_2	1.0868E+00	-9.3994E-03	8.7194E-03	-5.5426E-03	-9.9280E-03	-3.4565E-08	3.1892E-02	-1.7972E-03
a_{22}	-7.0697E-01	-1.4445E-03	-2.8507E-04	-1.7884E-03	-2.9672E-03	-2.3119E+04	-7.5402E-03	-2.0129E-03
a_{23}	-3.2902E-03	6.0609E-05	-6.3354E-06	-1.5890E-03	-6.0653E-06	1.1585E-28	-1.8452E-02	-7.3136E-03
a_{24}	1.1278E-01	6.6627E-04	5.1334E-05	4.8214E-04	1.3145E-03	-8.2118E-13	3.4533E-03	2.2887E-03
a_{25}	-1.5042E-02	-1.1317E-02	7.9945E-03	5.2906E-03	-4.5444E-03	-9.3013E-26	-4.4410E-02	-3.0728E-01
a_{26}	1.2744E-03	2.0832E-06	4.6923E-06	-1.0962E-12	8.5603E-06	-8.6891E-30	-4.1533E-05	1.3288E-05
a_{27}	1.2483E-04	1.1208E-05	7.8209E-07	1.4667E-05	2.7993E-06	-2.9428E-30	5.5870E-05	1.5195E-05
a_3	2.0549E-02	4.6545E-02	8.5685E-03	-9.8221E-02	5.7203E-03	1.5633E-03	-3.0741E-01	1.5279E-01
a_{33}	-9.4852E-03	-5.1758E-03	-5.2099E-04	-9.5787E-01	-5.3336E-05	-4.9061E-05	-7.0809E+00	-7.6644E+03

[cont.]

$\mathbf{a_{34}}$	-7.9576E-04	-1.0407E-04	-1.1195E-06	7.9705E-04	-1.2998E-05	1.6154E-13	5.1534E-03	-7.6623E+03
$\mathbf{a_{35}}$	2.5332E-02	3.6951E-02	5.6067E-03	8.3002E-02	6.6755E-03	1.9638E-02	1.1675E-01	1.5325E+06
$\mathbf{a_{36}}$	-6.2534E-06	-1.6195E-05	-5.6396E-06	-2.8920E-11	-1.4412E-06	-3.5942E-06	1.0221E-04	8.2718E-05
$\mathbf{a_{37}}$	5.8260E-04	-1.1162E-05	-5.3628E-08	-9.1489E-06	2.3836E-06	1.1016E-06	1.5964E-03	-6.9007E-05
$\mathbf{a_{4}}$	-1.0272E-01	2.6910E-03	-2.9591E-03	-1.5849E-02	-1.5647E-03	-1.5630E-08	-1.9480E-02	-3.2082E-02
$\mathbf{a_{44}}$	-4.4985E-02	-4.3043E-04	-5.0715E-05	-2.4413E-04	-8.0869E-04	-3.6192E+03	-2.1239E-03	-7.6629E+03
$\mathbf{a_{45}}$	3.9965E-01	1.2468E-02	8.5124E-03	1.7517E-02	1.9473E-02	-1.2960E-10	4.5902E-02	1.5326E+06
$\mathbf{a_{46}}$	8.1947E-05	1.0690E-05	-4.6069E-07	1.3716E-12	1.1117E-05	-5.0527E-13	4.0769E-05	-2.0050E-05
$\mathbf{a_{47}}$	8.0516E-04	4.2907E-06	1.7198E-06	4.9354E-06	9.6280E-06	-2.5940E-14	1.9471E-05	2.1789E-05
$\mathbf{a_{5}}$	4.2314E+00	3.7560E+00	4.4892E+00	6.9334E+00	3.2369E+00	3.5315E+00	5.6469E+00	1.4507E+01
$\mathbf{a_{55}}$	-1.8034E+01	-4.7213E+00	-7.4315E+00	-7.9169E+00	-7.3698E+00	-1.9069E+01	-5.0345E+00	-3.0652E+08
$\mathbf{a_{56}}$	-1.6811E-03	6.5263E-04	9.0237E-04	-1.3772E-10	1.0388E-03	1.5305E-03	-7.5767E-04	6.4756E-03
$\mathbf{a_{57}}$	-4.2077E-02	-6.9298E-04	-1.4800E-04	-1.1159E-03	-6.7674E-04	-6.9687E-04	-4.1437E-03	-2.7996E-03
$\mathbf{a_{6}}$	4.5617E-04	2.1649E-03	3.2037E-03	4.4546E-09	2.2925E-03	2.4997E-03	4.4986E-03	5.3594E-03
$\mathbf{a_{66}}$	-1.9384E-05	-3.1892E-06	-4.6882E-06	-1.9755E+02	-3.9905E-06	-2.6655E-05	-5.1975E-06	-5.3102E-06
$\mathbf{a_{67}}$	-1.9773E-05	-6.1822E-07	-4.8181E-07	-6.1887E-14	-1.0615E-06	-1.2332E-06	1.4544E-06	-6.3959E-07
$\mathbf{a_{7}}$	5.8905E-02	3.7966E-03	1.7629E-03	1.9985E-03	1.3859E-03	1.6362E-03	1.1816E-02	4.6793E-04
$\mathbf{a_{77}}$	-5.6784E-03	-5.6570E-06	-4.8357E-07	-1.1507E-06	-2.2614E-06	-6.2330E-06	-1.1574E-04	-1.2350E-05

TABLE A.4

Coefficients Used in the Set of Quadratic Discriminant Functions for Loads on Pavetta Street

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
$\mathbf{a_{0}}$	4.29006E+01	1.48361E+01	-3.61167E+00	2.84975E+01	2.04007E+01	1.99199E+08
$\mathbf{a_{1}}$	-1.18197E+01	1.14782E+01	8.16243E+00	-1.87833E+01	-2.29323E+01	-5.41465E+08
$\mathbf{a_{11}}$	-1.94919E+01	-1.60456E+01	-4.06139E+01	-7.27973E+00	-4.83936E+00	-2.39748E+07
$\mathbf{a_{12}}$	-8.62358E+00	-9.99360E+00	1.68178E+01	1.18618E+00	1.08300E+00	1.26883E+07
$\mathbf{a_{13}}$	6.23336E+00	3.60917E+00	2.99395E+01	2.95211E+00	4.15279E+00	3.17675E+06
$\mathbf{a_{14}}$	-6.43678E-01	8.08276E+00	-6.13792E+01	-5.59283E+00	-1.96336E-01	6.18550E+07
$\mathbf{a_{15}}$	8.57193E+00	1.19714E+01	2.99826E+01	1.82861E+00	2.52341E+00	-4.25038E+06
$\mathbf{a_{16}}$	3.83984E+00	1.40923E+01	1.28007E+01	-9.16390E+00	7.55776E-01	4.88302E+07
$\mathbf{a_{17}}$	-5.19224E+00	2.46178E+00	-2.94150E+01	-2.56322E+00	-2.39924E+00	3.19354E+07
$\mathbf{a_{2}}$	2.16853E+01	7.36998E+00	3.15032E+01	-5.41880E+00	-7.99613E+00	-3.25855E+08
$\mathbf{a_{22}}$	-1.70516E+01	-4.71431E+01	-6.17567E+01	-7.43177E+00	-9.56405E+00	-1.81576E+08
$\mathbf{a_{23}}$	-1.29966E+01	-4.23741E+01	-7.06750E+01	-1.41565E+00	-6.03618E+00	7.32789E+07
$\mathbf{a_{24}}$	-1.59797E+00	9.48645E+00	4.96252E+01	2.56894E-01	1.58109E+00	1.53254E+08
$\mathbf{a_{25}}$	-3.71270E+00	-2.81517E+00	-3.07754E+01	2.50434E+00	-3.05587E+00	2.27288E+07
$\mathbf{a_{26}}$	5.22835E+00	4.30586E+01	-1.50883E+01	-7.47118E-01	1.30567E+00	-3.78158E+07
$\mathbf{a_{27}}$	4.36173E+00	2.64732E+01	8.21660E+01	-1.75835E+00	3.91633E+00	8.06387E+07

[cont.]

$\mathbf{a_{3}}$	-9.88024E+00	-5.57789E+00	-4.71102E+01	-1.02190E+01	-1.60034E+01	-1.81455E+08
$\mathbf{a_{33}}$	-4.33896E+01	-1.11952E+02	-4.27367E+02	-1.09172E+01	-1.82277E+01	-2.99961E+08
$\mathbf{a_{34}}$	-1.69199E+00	-2.43777E+01	8.41260E+01	-1.25210E+00	-1.81864E-01	-1.37367E+08
$\mathbf{a_{35}}$	-2.04907E+01	-4.23317E+01	-2.47069E+02	-5.18134E+00	-7.85090E+00	1.28989E+08
$\mathbf{a_{36}}$	6.52815E+00	9.77013E+00	-3.89210E+01	1.47696E+00	-2.80579E+00	-2.58926E+07
$\mathbf{a_{37}}$	2.14657E+01	3.95874E+01	1.99340E+02	3.64286E+00	8.29927E+00	2.32218E+08
$\mathbf{a_{4}}$	8.62823E-01	-1.67085E+00	-1.79514E+01	2.25139E+01	2.84594E+00	-3.25611E+08
$\mathbf{a_{44}}$	-4.43368E+01	-1.12622E+02	-2.23377E+02	-5.50115E+01	-1.23281E+01	-3.66553E+08
$\mathbf{a_{45}}$	-2.78983E+00	-3.65022E+01	3.98712E+01	5.03957E-01	-2.83764E+00	1.45508E+07
$\mathbf{a_{46}}$	-1.77519E+01	-5.29988E+01	3.33645E+01	-2.51531E+01	-4.74766E+00	-1.17787E+08
$\mathbf{a_{47}}$	-4.71878E+00	-2.76978E+01	-9.99003E+01	-8.12589E-01	2.10264E-02	-1.34694E+08
$\mathbf{a_{5}}$	-1.08686E+01	-1.88945E+00	-3.31745E+01	-1.81397E+00	1.63086E+01	1.89178E+08
$\mathbf{a_{55}}$	-2.14169E+01	-5.43425E+01	-2.22651E+02	-7.58640E+00	-1.97634E+01	-7.35519E+07
$\mathbf{a_{56}}$	-3.27913E+00	-1.68974E+01	-7.70491E+00	3.55939E-01	-4.01455E+00	2.26583E+07
$\mathbf{a_{57}}$	8.53479E+00	-1.23690E+01	7.68831E+01	3.20610E+00	1.03201E+01	-1.48538E+08
$\mathbf{a_{6}}$	-8.01834E+00	-5.67928E+00	1.21981E+01	4.09088E+01	-1.30666E+00	-4.43724E+08
$\mathbf{a_{66}}$	-3.80397E+01	-1.14057E+02	-5.51307E+01	-3.27031E+01	-1.38645E+01	-1.02529E+08
$\mathbf{a_{67}}$	-6.42005E+00	-2.93979E+01	4.87611E+01	-5.32786E+00	-3.06929E+00	-3.28129E+07
$\mathbf{a_{7}}$	9.91884E+00	3.70871E+00	-1.99078E+01	9.41731E+00	-1.64002E+00	1.17308E+08
$\mathbf{a_{77}}$	-3.23922E+01	-1.23307E+02	-2.66177E+02	-1.33317E+01	-2.03328E+01	-3.86510E+08